Foreign Sourcing and Concentration in the U.S. Retail Sector

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Dominic Smith

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Thomas Holmes, Advisor

July, 2019

Acknowledgements

I am grateful to my thesis advisors Thomas Holmes, Teresa Fort, Amil Petrin, and Joel Waldfogel for their guidance. I would also like to thank Sergio Ocampo, the participants of the Applied Micro Seminar at the University of Minnesota, Center for Economic Studies, Federal Reserve Board, RDC Conference, and IIOC for helpful comments and suggestions. This paper is based upon work supported by the Doctoral Dissertation Fellowship at the University of Minnesota. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

Dedication

To my Fish. Although you will never read this, it would not exist without you.

Abstract

The U.S. retail sector has changed over the past three decades from one with many small firms to one dominated by large firms. The first chapter uses new data to document these changes at both the national and local level. It shows that concentration has been increasing in the U.S. retail sector at both the local and national level for the past 30 years. It shows that the primary contributor to the rise in national retailers is the growth of national retail firms that have stores in many markets and the exit of small retail firms. The second chapter of this work relates those changes to trade by establishing that stores of small retail firms are more likely to close when their competitors import directly. This work combines detailed measures of store sales with data on imports of each store's competitors to establish that stores of small firms are more likely to close when their competitors import directly. The final chapter of this work estimates a model of retailer entry decisions with direct imports to estimate the net effect of imports on local retail markets. I estimate that shutting down trade would not decrease local concentration because although small retail firms would enter markets, the largest retailers would exit.

Simultaneously, foreign sourcing of consumer goods has increased substantially, with much of that increase driven by large retailers' imports from China. This study examines the role of direct imports from China in the transformation of the U.S. retail sector. I propose two changes to measuring concentration. Existing work on concentration tends to study its evolution using national, industry-level data, but these metrics provide an incomplete picture given the local nature of competition in retail and the growing importance of multi-product general merchandisers who compete across industries. I therefore construct new data on store-level revenue for all U.S. retailers by 20 major categories of goods. While the national product-level Herfindahl-Hirschman Index more than doubled between 1997 and 2007, local concentration increased by only 50 percent. The new local-by-product concentration measures also enable me to perform an analysis of the role of globalization in increased concentration. I construct a measure of each small store's exposure to direct imports of large retailers. Using a store-level Bartik instrument (1991), the results suggest that a one percentage point increase in exposure to direct imports leads to a 0.7-1.7 percentage point increase in the probability a small store exits. I use a dynamic, continuous-time entry model to estimate the net effect of imports on the structure of competition in clothing sales, a product category highly exposed to direct imports. The results indicate that direct imports account for at least 14 percent of the decrease in the number of small clothing stores.

Contents

List of Tables

List of Figures

Chapter 1

The Evolution of U.S. Retail Concentration

1.1 Introduction

In the past 30 years, U.S. retailing has become substantially more concentrated. Between 1997 and 2007, the share of sales going to the 20 largest firms increased from 18.5 percent to 25.4 (Hortaçsu and Syverson, 2015). During this time period, the national Herfindahl-Hirschman Index (HHI) in retail doubled (Autor, Dorn, Katz, Patterson and Van Reenen, 2017). These patterns appear to be part of an economy-wide trend toward greater ownership concentration (Autor et al., 2017) and an increase in the dominance of large, established firms (Decker, Haltiwanger, Jarmin and Miranda, 2014). There is evidence that increases in concentration are accompanied by steeply rising variable markups (De Loecker and Eeckhout, 2017; Hall, 2018), which raises concerns about rising market power and prices.

Yet, most of the currently available evidence on concentration of the retail sector relies on national industry-based measures. These measures suffer from two major drawbacks. First, they describe national trends, although most retail markets are local, with consumers choosing between establishments in their geographical vicinity that sell a given product, making local concentration the relevant measure of the competitive environment in retail. The growth in national concentration can, but need not imply, concentration in local markets. To see this clearly, suppose that initially each U.S. city has a different largest store. Then suppose that a national retailer opens a store in each market, replacing the largest store but without displacing any business from the smaller stores. Then national concentration would rise, while local concentration would not. Alternatively, growth in national retailers might displace not just the largest stores but also smaller local ones, in which case growth in national concentration would be accompanied by a growth in local concentration. Whether the national expansion of large retailers, such as Walmart and Target, increases local retail concentration depends on whether they displace large, medium-sized, or small local retailers.

The second drawback is the reliance on industry-based measures, which do not account for the increasing importance of general merchandise establishments. Establishments in general merchandising compete with stores in many other industries. For example, Walmart is in the general merchandising subsector (3-digit NAICS 452) but competes with grocery, clothing, and toy stores.[1] Competition across industries can be important, as in the case of sales of grocery stores, where between 1992 and 2002 the share of grocery sales accounted for by general merchandisers increased from 6.9 to 18.2 (Basker and Noel, 2009).

This paper addresses both concerns by first providing a novel decomposition of the national HHI into a component driven by local market concentration and a new component that we call "cross-market" concentration which is driven by consumers in different markets shopping at the same firms. This decomposition is useful for separating the contribution to national concentration of changes in local markets from the expansion of large national retail chains (such as Walmart and Target) that took place during the last three decades. We show that the increase in national concentration is almost completely due to increases in the cross-market term. The second step in addressing the concerns raised above consists of measuring retail concentration using new data on store-level revenue for all U.S. retailers in 20 major categories of goods between 1982 and 2012. We combine two sources of confidential U.S. Census Bureau microdata that cover 1982 to 2012, namely the Census of Retail Trade and the Longitudinal Business Data. These new data allow us to measure concentration at the local level for each product category, taking into account the role of general merchandisers in each market. The coverage of the data makes it possible to document the evolution of U.S. retail concentration.

Using these data we document several new facts on concentration in the retail sector. First, we show that both the national and local HHI increase, but at different rates, with the national HHI increasing faster than the local HHI, particularly after 1997. The decomposition of national HHI into local and cross-market concentration, shows that the increases in national and local concentration are due to different factors. Changes in national concentration are driven by consumers in different markets shopping at the same stores (cross-market concentration), not changes in local concentration. The decomposition shows that

[1] References to specific firms are based on public data and do not imply the company is present in the confidential micro data.

98 percent of the change in national concentration is explained by changes in cross-market concentration. Second, we show that the increases in local concentration are broad-based. Most markets and product categories feature increasing concentration. The HHI increased in 81 percent of commuting zones accounting for 72 percent of retail sales in 2007. The HHI also increased in 7 of the major 8 product categories in retail. Finally, we document differences between industry-based and product-based measures of concentration. First, these measures are conceptually different, as they have different definitions of a market. These differences are increasingly important due to the rise of general merchandisers. Second, we show that product-based measures of concentration (that incorporate competition from general merchandisers) are lower than industry-based measures. The average product-HHI at the national and local level is about a third of the average industry-HHI. However, all measures of concentration share an upward trend.

Our main contribution is to the growing literature on concentration. As is well understood by researchers, there is an important distinction between national and local concentration in retail, thus, we contribute by providing new measurements of concentration for local markets and relating them to the increase in national concentration and the growing importance of multi-market firms. The closest paper to ours is Rossi-Hansberg, Sarte and Trachter (2019), which evaluates changes in concentration at both the national and local levels in multiple sectors (e.g. manufacturing, retail, etc.). Using the U.S. National Establishment Time Series (NETS) establishment-level data set, they find that between 1992 and 2012, concentration at the national level increased in six major sectors, while local concentration decreased.[2] The findings in this study at the national level are similar to those in Rossi-Hansberg et al. (2019). However, our results at the local level differ sharply, as we find concentration at the local level has generally been increasing across the various product categories in the retail industry.[3] There are multiple reasons for our results to differ, but the major difference is the data source. Data in the present study are based on confidential information collected by the Census Bureau and Internal Revenue Service. They are considered the gold standard for measuring economic activity at the store level. These records make it clear that local concentration has been increasing, though not as much as at the national level. More details on these differences are provided in Appendix A.1.

[2]For the retail sector national concentration increased by five percentage points, while local level concentration decreased by 14 percentage points. Numbers are taken from the retail sector line in Figure 1b in Rossi-Hansberg et al. (2019).

[3]These results are in line with other studies that have documented that local trends in retail may differ from local trends in other sectors. In particular, Rinz (2018) and Lipsius (2018) look at concentration in labor markets using census micro-data, finding increasing labor market concentration in retail, but decreasing labor market concentration overall.

This study contribute to the literature on concentration by measuring national and local concentration using product-level revenue, which handles the multi-product nature of large retailers. There has been substantial work documenting increasing national retail concentration at the industry level (Foster, Haltiwanger, Klimek, Krizan and Ohlmacher (2015); Hortaçsu and Syverson (2015); Autor et al. (2017)), but much less work measuring concentration at the local level (Rossi-Hansberg et al., 2019). Results show that national and local concentration are increasing at both the product and industry level.

More broadly, this paper contributes to the growing body of work on increased national concentration, and its relation to the declining labor share (Autor et al., 2017), and the declining churn and reallocation of aggregate activity to large established firms (Decker et al., 2014). These trends may reflect increased allocative efficiency but also raise concerns about market power and rising prices (De Loecker and Eeckhout, 2017). Despite a broad consensus on increased concentration, there is little evidence on the mechanisms driving the change. This paper provides a new result linking changes in local concentration and expansion of firms across locations to national concentration. We apply this result to new micro-data focusing on a specific sector, retail, in which the growth in aggregate concentration has been particularly dramatic. The results make it clear that increases in national concentration need not imply increasing local concentration, making it important to distinguish between local and aggregate trends when assessing the effects of concentration in the economy.

Finally, the decomposition of the HHI presented in this paper contributes to the literature on the measurement of concentration by developing a new interpretation of the HHI. Multiple decompositions of the HHI have been developed, including a decomposition into a component due to the number of stores in a market and a component due to how unequal the market shares of the stores are. We interpret the HHI as a probability, namely the probability that two dollars spent at random in a market are spent in the same firm. This interpretation allows for multiple decompositions based on the Law of Total Probability. We focus on decomposing the national HHI into an average of local HHIs and a cross market term that captures consumers in different markets buying goods at the same firm. However, this approach can also be used to bound the potential error in the HHI based on market mispecification among other applications.

The rest of the paper proceeds as follows. Section 1.2 establishes the relation between local and national concentration and develops a decomposition of national concentration into local and cross-market concentration. Section 3.2 describes the data, including how to construct store-level sales by product. Section 1.4 measures national and local concentration and establishes the main facts about their evolution since 1982. Section 1.5 describes how the results vary across locations and products. Section 1.6 shows that defining retail markets using industry data can be problematic. The final section concludes.

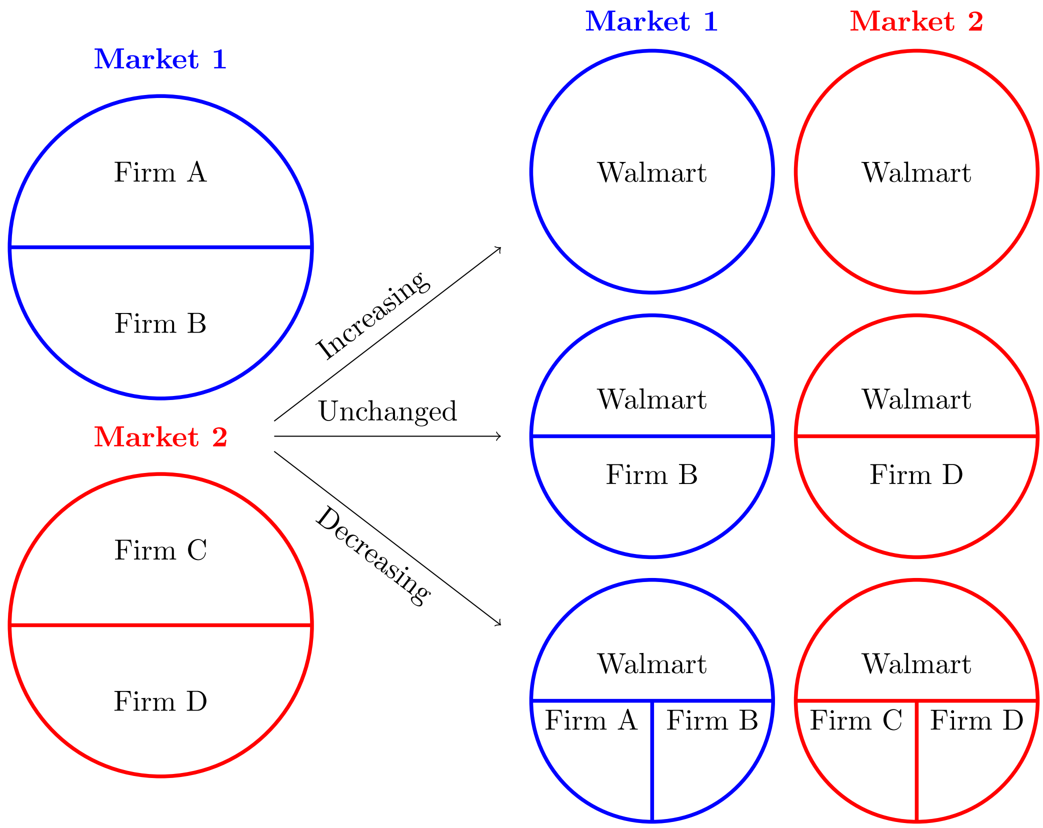
1.2 National and Local Concentration

The increasing trend of national concentration in various sectors of the economy has been widely documented, and has attracted the attention of the literature, in part due to the potential impacts of concentration on competitive behavior, firm growth, declining labor share, innovation, and other aspects of firm behavior (i.a., Autor et al., 2017; Decker et al., 2017; Akcigit and Ates, 2019). One major outstanding question in this literature concerns the mechanisms behind the increase in national concentration. In particular, what is the role of local concentration in the rise of national concentration? Consumers buy many goods and services in their local area, making local markets relevant for competition (Rossi-Hansberg et al., 2019).

Increasing national concentration can be accompanied by increasing local concentration, but it may also be accompanied by decreases in local concentration. In fact, not much can be learned from the dynamics of local concentration having information only about national trends. The simple example shown in Figure 1.1 makes this clear. National concentration can increase by having firms expand across markets, without affecting the layout of individual markets (row 2). Alternatively, the expansion of large firms can drive out competitors in local markets, increasing national and local concentration (row 1), or can bring up more—and likely smaller—competitors, in turn decreasing local concentration (row 3). The total effect on national and local concentration depends on how firms in individual markets respond.

The example in Figure 1.1 highlights the two mechanisms affecting national concentration that we study in this paper. National concentration is affected by changes in local concentration and in cross-market concentration. The first mechanism draws the link between changes in the layout of local markets and concentration at the national level. As local markets become more/less concentrated so does the aggregate economy (in a given sector). The second mechanism links national concentration to cross-market concentration, that is, the presence of the same firms across various markets. As firms expand across markets they capture a larger share of national sales, in turn increasing national concentration. Note that, as shown in Figure 1.1, changes in cross-market concentration need not be accompanied by changes in local concentration. In what follows we make these ideas precise by developing a new decomposition of national concentration into local and cross-market concentration.

Figure 1.1: Effect of Increasing National Concentration on Local Concentration



Notes: Figure shows hypothetical market structures after the entry of Walmart into markets.

We measure concentration with the Herfindahl-Hirschman Index (HHI). The HHI is one of the most common measures of concentration and its formulation will prove useful in decomposing the mechanisms behind the changes in national concentration. We measure concentration at the product category level throughout the paper,[4] using each firm's, i, share of sales in product j at time t, s_i^{jt}.[5] The national HHI in a year is defined as the sum of the product-level HHIs in each year, weighted by the share of product j's sales in total retail sales, s_j^t:

$$HHI^t = \sum_{j=1}^{J} s_j^t \sum_{i=1}^{N} \left(s_i^{jt}\right)^2, \tag{1.1}$$

[4] Distinguishing markets by product categories is critical to determine which establishments (or firms at the national level) are in direct competition with one-another. This is particularly relevant in the retail sector as discussed in Section 3.2. Details of the definition of product categories are presented in Section 3.2 and in Appendix B.1.2.

[5] Superscripts and subscripts are defined such that s_a^b is the share OF a IN b.

while the local HHI of market m and product j in year t is is calculated as:

$$HHI_{mj}^{t} = \sum_{i=1}^{N} \left(s_i^{jmt} \right)^2 . \tag{1.2}$$

The formulation of the HHI is useful in understanding the connection between local and national concentration. From (1.1) it is clear that the national HHI for a product j measures the probability that two dollars, x and y, chosen at random, are spent at the same firm.[6] This interpretation of the HHI as a probability is what allows us to capture the separate role of local and cross-market concentration. Intuitively, the probability that two dollars are spent in the same firm goes up if the firm has captured more of the sales in the market(s) it is present (increase in local concentration), or if the firm is present in more markets (cross-market concentration).

The interpretation of the HHI goes beyond providing intuition for the mechanisms affecting national concentration. Since the index is a probability, it can be decomposed using the law of total probability into two terms: a weighted average of local concentration and a residual term, which we refer to as cross-market concentration. As mentioned above, Local concentration measures the extent to which consumers in a local market shop at the same firm, while cross-market concentration measures the degree to which consumers in different locations shop at the same firm. The decomposition is given by:

$$\underbrace{P(i_x = i_y)}_{\text{National HHI}} = \overbrace{P(m_x = m_y)}^{\text{Collocation}} \underbrace{P(i_x = i_y | m_x = m_y)}_{\text{Local HHI}} + \overbrace{P(m_x \neq m_y)}^{\text{1 - Collocation}} \underbrace{P(i_x = i_y | m_x \neq m_y)}_{\text{Cross Market HHI}}, \tag{1.3}$$

where i_x is the firm at which dollar x is spent and m_x is the market in which dollar x is spent, likewise for y.[7]

Equation (1.3) has three components. The first component, $P(m_x = m_y)$, which we term collocation, is the probability that two dollars are spent in the same market. The collocation term is given by:

$$P(m_x = m_y) = \sum_{m=1}^{M} \left(s_m \right)^2 , \tag{1.4}$$

where s_m is the share of market m in national sales. The second component, $P(i_x = i_y | m_x =$

[6] In what follows, the j and t superscripts are dropped on all variables for convenience.

[7] It is worthwhile noting that this is not the only decomposition that can be obtained using this method. The particular decomposition in (1.3) is motivated by the question on the connection between national and local trends in concentration. Appendix A.2 presents further details on the decomposition.

m_y), is an aggregate index of local concentration, with local concentration measured as in equation (1.2).[8] The third component, $P(i_x = i_y | m_x \neq m_y)$, which we call cross-market concentration, captures the probability that a dollar spent in different markets is spent at the same firm:

$$P(i_x = i_y | m_x \neq m_y) = \underbrace{\sum_m \sum_{n \neq m} \frac{s_m s_n}{1 - \sum_p s_p^2}}_{\text{Weights}} \underbrace{\sum_{i=1}^{N} s_i^m s_i^n}_{\text{Cross Market}} \tag{1.5}$$

The cross-market concentration index between two markets (say m and n) is given by the product of the shares of the firms in each market (the probability that two dollars spent one in each market are spent in the same firm). The pairs of markets are then weighted by their share of sales and summed.

Of the three terms, the collocation term plays a crucial role in determining the extent of the impact of local concentration in national measures. A low collocation term implies that local concentration can only have a limited effect on national trends, leaving the cross-market term as the driver of the national index. We will show later that this is in fact the case.[9]

To implement the decomposition presented in equation (1.3) we need to measure concentration in each local market for a given product, as well as to link the activities of firms across markets. Doing this requires detailed data on establishment-level revenue by product for all firms in the U.S.. In the next Section we use confidential microdata from the U.S. Census Bureau to construct a new data set that allows us to measure local concentration and implement the decomposition. We find that local concentration increases in the U.S., albeit at a slower rate than national concentration. Nevertheless, the decomposition shows that national concentration is almost entirely determined by cross-market concentration, with the increase in local concentration having a very limited impact on the increase of national concentration. As explained above, this is due to the low value of the collocation term in the data. Sections 1.4 and 1.5 expand on these results.

[8] In the decomposition each local market is weighted by the conditional probability that the two dollars are spent in market m, given that they are spent in the same market: $s_m^2/(1-\sum_p s_p^2)$. This tends to weight larger markets more than the more usual weight s_m—the share of sales (of product j) accounted by market m (at time t). To facilitate comparison of the results with other work, we present aggregated series for local concentration in Section 1.4 that use the latter weights.

[9] The low value of the collocation term should come as no surprise, because the U.S. has many markets and even the largest markets represent only a small fraction of total U.S. sales.

1.3 Data: Retailer Revenue for All U.S. Stores

This section describes the creation of new data on store-level revenue for 20 product categories for all stores with at least one employee in the U.S. retail sector. These data allow for the construction of detailed measures of concentration that take into account competition between stores selling similar products in specific geographical areas.[10]

1.3.1 Data Description

This paper combines two sources of confidential U.S. Census Bureau microdata that cover 1982 to 2012. The primary source of data is the Census of Retail Trade (CRT), which provides revenue by product type for retail stores in years ending in 2 and 7. The CRT data on product-level revenue and information on the location of each store are used to define which stores compete with each other. Importantly, a store's local competition will include stores in many different industries inside the retail sector, because stores of different industries can sell similar products. This is particularly evident for stores in the General Merchandising sub-sector, composed by establishments selling multiple product types. The data we create here are uniquely equipped to deal with cross-industry competition.

The CRT is combined with the Longitudinal Business Data (LBD) to identify the activity of stores of each firm in other sectors of the economy. This information assists in tracking the behavior of firms with a presence in retail in manufacturing, wholesaling, and other sectors of the economy.

1.3.2 Sample Construction

The **retail sector** is defined based on the North American Industrial Classification System (NAICS) as stores with a 2-digit NAICS code of 44 or 45. As such it includes stores that sell final goods to consumers without performing any transformation of materials. For 1992 to 2012 we use the 8-digit industry codes (NAICS) available from the CRT as the industry of store. Prior to 1992 we use the 6-digit industry codes (NAICS) from Fort and Klimek (2016).

The sample includes all stores with at least one employee assigned to the retail sector. The number of stores increases over time, while the number of firms decreases. Employment increases over time representing about 10 percent of U.S. employment over the whole sample period. The retail sector is estimated to account for about 5.5 percent of GDP by the BEA in 2007.

[10] We use store and establishment as synonyms.

1.3.3 Creation of Department-Level Revenue

The CRT asks establishments to provide data on revenue by product line (for example, men's footwear, women's pants, diamond jewelry). Unlike in other sectors of the economy, retail stores compete with stores in other industries. In particular, general merchandise stores such as Walmart and Target, compete with stores in clothing, groceries, and electronics. By 2007, general merchandisers accounted for a significant fraction of sales in many of these departments. For example, general merchandisers accounted for 44 percent of clothing sales in 2007 (CRT, 2007). Thus, revenue by product line is important when looking at competition in the retail sector. The product line codes are aggregated into 20 departments such that stores in industries outside of general merchandise sell primarily in one department. For instance, stores in subsector 448 (clothing and clothing accessory stores) primarily report sales in products such as women's dress pants, men's suits, and footwear, which are grouped into a clothing department. Table B.1.2 lists these departments.

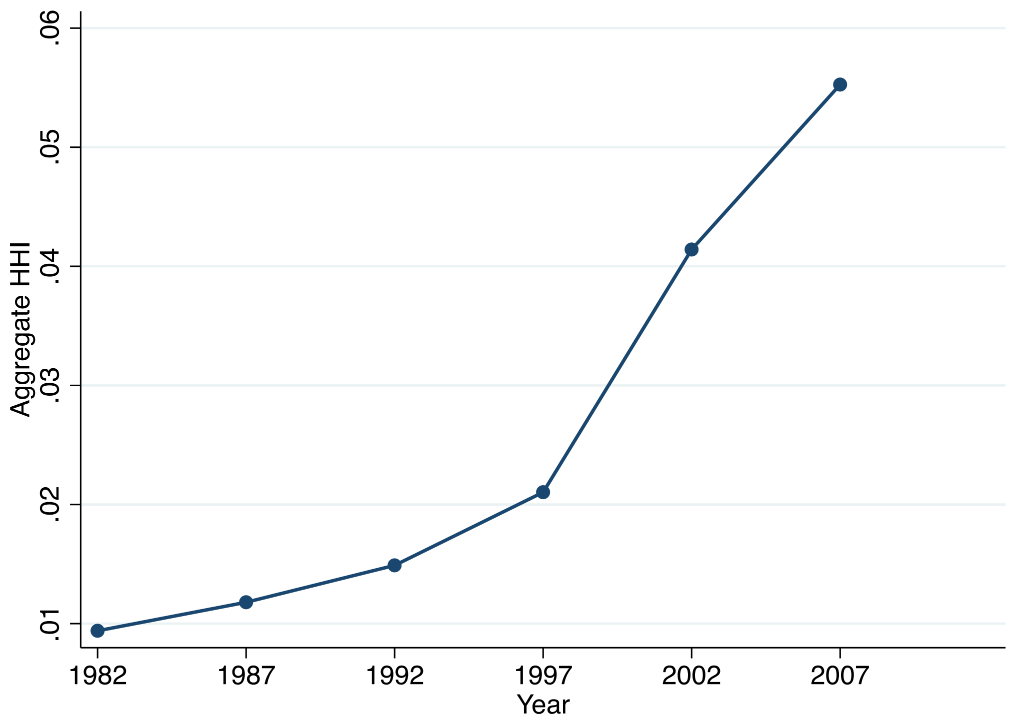
Aggregating data in this allows for accurately imputing revenue by department for stores that do not report product line data. The CRT only asks for detailed product lines from a sample of small stores. For the remainder, store-level revenue estimates are constructed from administrative data, without revenue by product line. This affects stores that account for 20 percent of sales. For these stores, the distribution of their sales across departments are imputed using characteristics of the store, such as industry and multi-unit status. Details of this procedure are provided in Appendix B.1.1.

1.4 Changes in Retail Concentration

This section exploits the detailed micro data described in Section 3.2 to decompose national concentration in the U.S. retail sector into local and cross-market concentration using the identity developed in equation (1.3). In doing so, we calculate a new measure of concentration that accounts for the local nature of competition in retail and the rise of general merchandisers that compete with stores across multiple subsectors. This new measure has a distinct advantage in that it measures concentration at the product category level, rather than the industry level. We show that local concentration has increased, although not as much as national concentration. Moreover, the decomposition reveals that national concentration is largely independent of local trends, with 98 percent of the growth in national concentration accounted by increasing cross-market concentration (consumers shopping at the same firms across markets).

Figure 1.2 plots national concentration in the U.S. retail sector as measured by the HHI defined in equation (1.1). Between 1982 and 1997 national concentration was low, although it gradually increased over the period. In contrast, between 1997 and 2007, concentration grew at a faster pace, more than doubling from 0.02 to 0.055. These results are consistent with previous work documenting increasing national industry-level concentration in sales and employment across sectors, including retail.[11]

Figure 1.2: National Concentration

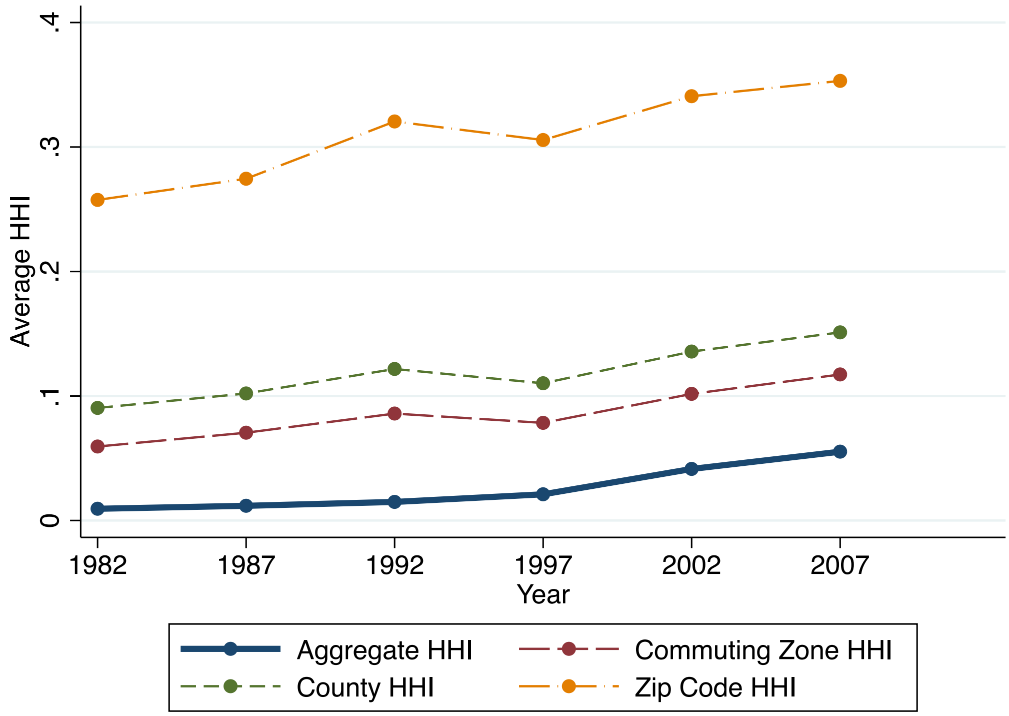


Notes: The data are from the CRT micro data set. Weighted averages of national HHI in eight major departments were computed.

Striking as it is, the increase in national concentration shows only part of the changes undergone by the retail sector during the past decades. Figure 1.2 alone does not provide information on the mechanisms underlying the change in national concentration. In particular, it does not account for what has happened to local retail markets. To determine the evolution of local concentration and its relation to national concentration we exploit the data constructed in Section 3.2 and implement the decomposition described in Section 1.2. Figure 1.3 plots the level of national and local concentration between 1982 and 2007. Local

[11]See Rossi-Hansberg, Sarte and Trachter (2019); Autor, Dorn, Katz, Patterson and Van Reenen (2017); Foster, Haltiwanger, Klimek, Krizan and Ohlmacher (2015); Basker, Klimek and Hoang Van (2012).

Figure 1.3: National and Local Concentration



Notes: The data are from the CRT micro data set. The eight main departments are weighted by sales share. The local HHI is aggregated using each market's share of national sales.

concentration increases whether markets are defined by zip codes, counties, or commuting zones, and are sustained throughout the sample period, with the exception of the mid 1990's.[12] Contrary to the national concentration index, local concentration did not accelerate its increase in the period after 1997. These results imply that if market power has been increasing, the increases are much more modest than those implied by national data.[13]
The picture that emerges from the data in the present study differs from the findings of Rossi-Hansberg, Sarte and Trachter (2019) (RST), who find that local retail concentration has been steadily falling since 1992. Our results differ for multiple reasons. First, a different data set is used.[14] Second, different definitions of which stores are retailers are employed.

[12]These results are consistent with studies on labor market concentration that find increasing concentration in retail, but decreasing concentration overall (Rinz, 2018; Lipsius, 2018).

[13]Local concentration may be correlated with market power, but local concentration measured by revenue can increase in response to large firms lowering their prices. Thus, the increases in local concentration may not imply decreases in welfare.

[14]RST use U.S. NETS data.

RST use Standard Industrial Classification (SIC) codes while this paper uses NAICS.[15] Finally, the aggregate index of local HHI is calculated differently. RST report the average change in the local HHI, weighting by the end-of-period sales/employment of each market, while we report the change in the average local HHI, weighting markets in each year according to that year's sales. This distinction matters because as markets become bigger, they also tend to become less concentrated. This mechanically gives more weight to markets where concentration is decreasing. In fact, when we repeat our exercise using end-of-period weights we find small decreases in local concentration when measured at the industry or product level. The results are presented in appendix A.1. We choose current period weights in order to be able to decompose national concentration as described in section 1.2. More details on differences between our studies are in appendix A.1.

Table 1.1: Collocation Term by Year

Department	1997	2007
Furniture	0.012	0.013
Electronics and Appliances	0.015	0.013
Home and Garden	0.009	0.009
Groceries	0.012	0.012
Health Goods	0.012	0.011
Clothing	0.015	0.016
Toys	0.011	0.009
Sporting Goods	0.013	0.016

Notes: The data are from the CRT micro data set. Collocation is the probability that two dollars chosen at random are spent in the same market. It measures the contribution of local concentration to national concentration. Markets are defined as commuting zones.

Having measured local concentration for each market and product category it is now possible to measure the relation between local and national concentration. We do this by means of the decomposition in equation (1.3), which breaks national concentration into local and cross-market concentration. The first result is the limited effect of local concentration on national concentration. The contribution of local concentration to national concentration is weighted by the collocation term—the probability that two dollars spent in the U.S. are spent in the same market, shown in Table 1.1 for the eight major product categories.

[15] The primary difference between SIC and NAICS is that SIC includes restaurants in retail.

Because the share of any given market in total U.S. retail sales is small (given the number of markets in the U.S.), the collocation term is quite small. This probability is less than 2 percent for all departments and all years, and has barely changed over time. As a result, a 10 percentage point increase in the local HHI will increase the national HHI by less than 0.2 percentage points. Moreover, the collocation term shows that in the extreme case where every market had only one firm, implying the local HHI is one, the national HHI would be less than two percent if each firm was only in one market.

The second result is the major role of cross-market concentration in shaping the national concentration index, this comes as the flip-side of the limited role of local concentration. National concentration has increased because consumers in different locations are shopping at the same (large) firms, in fact, 98 percent of the change in national concentration reported in Figure 1.2 is accounted for by changes in cross-market concentration. The probability that two dollars spent in the same product category are spent at the same firm in two different markets increased from around 2 percent to 5.5 percent in just 10 years (from 1997 to 2007). Put another way, product-level concentration increased from the level implied by 100 equal-sized firms to the number implied by 20 equal-sized firms. On the other hand, the probability two dollars spent in the same market are spent at the same firm increased from 7.8 percent to 11.7 percent, an increase of 50 percent.[16]

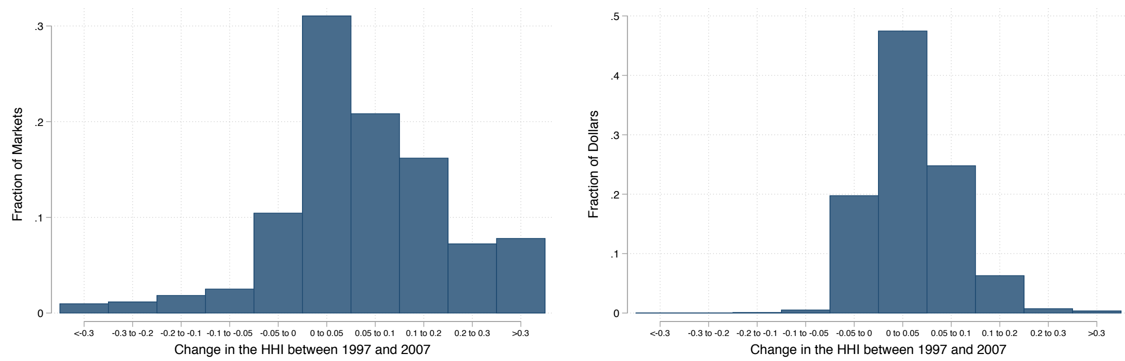
The Department of Justice merger guidelines consider a market to be highly concentrated if the HHI exceeds 0.25 and moderately concentrated if the HHI exceeds 0.15 (Justice and Commision, 2010). We find that the average retail market is just below the threshold for moderate concentration when markets are defined as a commuting zone. Markets measured by zip codes and counties are highly and moderately concentrated, respectively. We find similar results using the share of the top four firms in each product category. We find that in 1987 the top four firms in a commuting zone accounted for 4.4 percent of all sales in a department. By 1997 this share increased by 3.5 percentage points to 7.9 percent of sales. In the next 10 years the share of the top four firms grew by 7.2 percentage points to 15.1 percent.

These results show how markets have changed on average, but they miss large changes in the underlying distribution of concentration at the local level. The next section focuses on exactly those changes.

[16] Unless otherwise stated local concentration numbers refer to markets defined as commuting zones. Commuting zones are such that the majority of individuals work and live inside the same one. It seems likely that if individuals live and work in a commuting zone they do the majority of their shopping in that region.

Figure 1.4: Changes in Concentration Across Markets

(a) Unweighted (b) Weighted



Notes: The data are from the CRT micro data set. Left panel shows the fraction of markets, commuting zone/product category pairs, with changes in concentration of a given size. The right panel weights markets by value of sales in the department.

1.5 Variation in Local Concentration

This section delves deeper into the changes in local concentration experienced by markets in the U.S. during the last decades. As mentioned above, local concentration has risen on average since 1982, yet this general trend provides little information for how the change in concentration is distributed across markets. Using our detailed data on revenue by product for U.S. establishments we show that the increases in concentration were broad based across both products and locations.

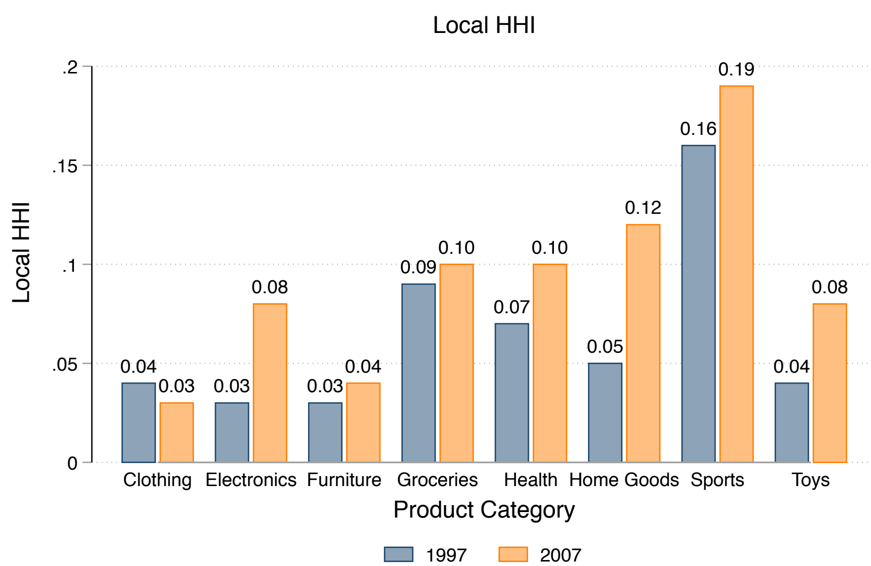
1.5.1 Changes in Concentration Across Locations

The increases in concentration have been very broad-based. Over 72 percent of dollars spent in 2007 are spent in markets which have increased concentration since 1992. Figure 1.4 shows the distribution of changes in concentration between 1997 and 2007. In just 10 years 52 percent of markets had increases in concentration over 5 percentage points and 32 percent of dollars were spent in markets with an increase of at least this size. These changes are significant. One criterion used by the Department of Justice to determine when to challenge mergers is whether the local HHI will increase by 2 percentage points, if the market is already highly concentrated (HHI over 25 percent).

1.5.2 Changes in Concentration Across Products

Changes in local concentration were also broad-based across products. Figure 1.5 shows that concentration increased for seven of the 8 major product categories. Although there is significant variation across products in terms of the level of the local HHI and the change. While some products, like electronics, home goods and toys, experience large increases in concentration, more than doubling the local HHI in the decade from 1997 to 2007, some others, like furniture and groceries barely increase.

Figure 1.5: Changes in Local Concentration Across Product Categories



Notes: The data are from the CRT micro data set. Numbers are the local HHI for various product categories weighted by market size.

1.6 Issues with Industry-Based Measures

A central contribution of this paper is the creation of store-level sales by product category for all U.S. retail stores. This allows us to define competition based on products rather than industry-based measures. Industries, either NAICS or SIC codes, are regularly used to define markets. This approach is often necessitated by data availability and in many sectors is likely to be a good approximation (e.g. manufacturing).

This is not the case in the retail sector. The retail sector has one set of industries, general merchandise stores (NAICS 452), that compete with stores in many industries. By construction these industries are composed by establishment that sell many types of products. Thus, industry-based measures ignore the competition faced by stores selling a given product, coming from general merchandise stores. The measures we developed in Section 1.4 overcome this shortcoming.

Table 1.2 presents industry-based and product-based concentration measures. There are two industry-based measures, the first one (NAICS-based) calculates concentration separately for all 6-digit industries in NAICS, while the second one (Selcect NAICS) calculates concentration for all 6-digit NAICS excluding auto dealers and auto-parts stores (441), gasoline stations (447), and non-store retailers (454). The product-based measure calculates concentration for the eight major product categories discussed in Section 1.5. As discussed above, each measure captures different concepts, as they define a market in a different way. These differences are more than just conceptual. The level of the different measures gives a different picture of how concentrated markets are. Product-based measures are about a third of the Select NAICS measure and half of the NAICS-based measure with all industries.[17] Despite their differences all measures of concentration exhibit similar dynamics, with national concentration measures increasing five- to six-fold since 1982, and local concentration measures roughly doubling.

[17]The differences in level across concentration measures cross the different thresholds for concentration establish by the Justice Department.

Table 1.2: Industry-Based and Product-Based Concentration Measures

	National Concentration						
	1982	1987	1992	1997	2002	2007	2012
NAICS-based	0.019	0.029	0.029	0.046	0.085	0.105	0.116
Select NAICS	0.030	0.043	0.046	0.080	0.143	0.182	0.195
Product-based	0.010	0.012	0.015	0.021	0.041	0.055	
	Local Concentration - Commuting zone						
	1982	1987	1992	1997	2002	2007	2012
NAICS-based	0.120	0.143	0.143	0.160	0.203	0.226	0.246
Select NAICS	0.155	0.184	0.191	0.222	0.279	0.313	0.326
Product-based	0.059	0.071	0.086	0.078	0.102	0.117	

Notes: All concentration mesures correspond to average Herfindahl-Hirschman indeces. NAICS-based measures are calculated including all NAICS industries. Select NAICS drops subsectors 441, 447, and 454. Product-based measures calculate concentration for eight major product categories. Product-based measures for commuting zones in 2012 have not been disclosed.

1.7 Conclusion

Despite the attention given to the rise of national concentration in the U.S., less is known about the dynamics of local concentration, and the relationship between observed national trends and the behavior of local markets. Knowing the state and dynamics of local concentration is of particular importance in the retail sector due to the local nature of most retail markets during most of the last three decades. This paper helps to shed light on these issues by contributing in two related fronts. First, we develop analytical tools that relate local and national concentration. Our tools allow us to decompose national concentration measures into a local component (national concentration rises as local markets become more concentrated), and a cross-market component (national concentration rises as the same firms are present in more markets, increasing their national market share). Second, we construct new data from confidential Census micro-data that allow us to measure concentration at a granular level. With this data we can define a market as a pair of a geographical location (e.g. a commuting zone), and a product. This data not only leads to new measures of product-based local concentration, but, in combination with the decomposition we propose, it also gives a clear picture of the mechanisms behind the increase in national concentration in the retail sector.

We show that local concentration has a limited effect on national concentration measures, even though concentration increases broadly across locations and products since the 1980's. Instead, it is cross-market concentration what explains most of the increase in national concentration observed since 1982. That is, national concentration is driven by consumers in different markets (locations) shopping at the same firms, highlighting the role of large multi-market retailers in explaining the dynamics of the retail market.

The unique data we construct also overcomes a drawback of previous measures of concentration, namely, the reliance in industry-based measures. By exploiting information on product-line sales we can define markets based on products. This makes it possible to capture competition across industries, a salient feature of retail markets due to the rise of general merchandise stores. This type of competition is ignored by industry-based measures, which cannot account for the competition faced by clothing or grocery stores, coming from general merchandisers. We show that this conceptual difference in the definition of markets (by industry or by product) lead to very different levels of concentration, with product-based measures being between a third and half of industry based measures, at both the national and the local level. However, all measures have similar upward trends, with national measures increasing by five- to six-fold and local measures roughly doubling.

Chapter 2

Foreign Sourcing and the U.S. Retail Sector

2.1 Introduction

In the past 30 years, U.S. retailing has become substantially more concentrated. Between 1997 and 2007, the share of sales going to the 20 largest firms increased from 18.5 percent to 25.4 (Hortaçsu and Syverson, 2015). During this same time period the number of stores of small retail firms decreased by seven percent. This pattern appears to be part of an economy-wide trend toward greater ownership concentration (Autor, Dorn, Katz, Patterson and Van Reenen, 2017) and an increase in the dominance of large, established firms (Decker, Haltiwanger, Jarmin and Miranda, 2014). These increases in concentration are accompanied by steeply rising variable markups (De Loecker and Eeckhout, 2017; Hall, 2018), which raise concerns about rising market power and prices.

Despite the consensus that large firms account for a growing share of activity, we lack an understanding of why this reallocation has occurred. One possibility is globalization. The growth of large U.S. retailers coincided with the period during which imports of consumer goods from China increased five-fold. Many of these consumer goods were imported by large retailers (Holmes and Singer, 2018) at significant marginal cost savings over smaller retailers that used intermediaries (Ganapati, 2018). While existing research finds that Chinese imports have led to lower prices and increased varieties for consumers, those imports disproportionately favor large U.S. retailers.[1] If Chinese imports have led to increased concentration in retail, this might mitigate—or eventually even reverse—the benefits of lower-priced consumer goods from China for U.S. consumers.

[1] See Amiti, Dai, Feenstra and Romalis (2017); Jaravel and Sager (2018); Handley and Limão (2017)

In this paper, the role of direct imports from China in the transformation of the retail sector is examined using data on store-level revenue for all U.S. retailers in 20 major categories of goods. The data allow for the construction of a store-level measure of exposure to direct imports based on the importing activity of a store's local competitors. I use these data to explore the causal impact of imports on the exit decisions of stores of small firms, documenting that increasing exposure to direct imports leads to the exit of small stores.

Trade can explain the exit of small stores because the fixed costs of trade are large (Antras, Fort and Tintelnot, 2017). Large retailers pay these fixed costs and buy much of what they sell directly from foreign suppliers, while small retailers pay higher costs to purchase through intermediaries. Large retailers, defined as firms with more than 100 stores, are responsible for more than 90 percent of total imports by retailers and have been linked to much of the growth in consumer goods imports from China (Basker and Van, 2010). Existing work has documented that competition with large retailers causes small retailers to exit.[2] Meinen and Raff (2018) show that small stores exit and local concentration increases in industries where retailers directly import more of the products they sell. Yet, these estimates rely on import exposure measures at the industry-level, which have the potential to miss important features of trade because some of the biggest importers are general merchandisers, who compete with stores in other industries.

In this study, detailed data on sales by product category are combined with data on firm imports by these same product categories. This allows for the construction of a measure of each store's exposure to trade based on which products each store sells and the actions of their local competitors. I test whether stores that become more exposed to direct imports between 2002 and 2007 are more likely to exit the market. This is accomplished while controlling for the size and scope of each store's competitors. An issue in this exercise is that increases in exposure to direct imports are correlated with increased competition from large firms. Increased competition from large firms may cause small stores to exit for reasons unrelated to direct imports. This fact can lead to biased estimates of the effect of direct imports on small store exit.

These concerns are addressed using an instrumental variables (IV) strategy similar to Hummels, Jørgensen, Munch and Xiang (2014). The instrument uses growth of China's exports by product category between 2002 and 2007. The IV identify stores that become more exposed to direct imports because of increases in exports from China in products already imported by their competitors in 2002. The change in import exposure of each store is calculated by assuming that product-level imports by each store's competitors grow at the

[2]See Jia (2008), Arcidiacono et al. (2016), Haltiwanger, Jarmin and Krizan (2010), Basker (2005)

rate of exports from China of that product. The instrument depends on the assumption that firms already importing in 2002 did not experience a simultaneous increase in competitiveness due to other factors in the same products in which exports from China grew. Results show that a one percentage point increase in the share of competitor's sales that are imported directly is associated with an increased exit rate of small stores by 0.7-1.7 percentage points.

The most similar paper to this one is work by Meinen and Raff (2018). They use data on Dutch retailers to show that retailers becoming importers is positively associated with higher sales, profits, and markups. This paper advances on their work in two key respects. First, I focus on the intensive margin of importing decisions, while Meinen and Raff (2018) focus on the effect of a firm becoming an importer. I show that the retail firms that begin importing between 1997 and 2007 are small, account for a small fraction of aggregate sales, and import a small fraction of their sales directly. Thus, I studying the impact of large retailers increasing their direct import penetration over the time period. Second, I measure each store's exposure to trade based on which product categories the store and its competitors sell. This is an improvement over traditional industry-based measures because it accounts for the fact that within firm there is significant variation in product category import penetration.

This work contributes to ongoing research on the role of foreign sourcing by studying how it impacts the decisions of retail firms. Previous work has shown that retailers play an important role in imports from China (Bernard, Jensen, Redding and Schott, 2010). Ganapati (2018) shows that intermediary markups can be substantial, which implies the marginal cost savings of importing directly can be large. The focus of this paper is on the domestic impact of direct imports by large firms on small retailers.

The rest of the paper proceeds as follows. Section 2.2 lays out a model of retailer importing. Section 2.3 describes the data, including how to construct store-level sales by product and import exposure. Section 2.4 provides descriptive evidence on the importance of importing to the retail sector. Section 2.5 provides estimates of the impact of direct imports on entry, exit, and growth in the retail sector. The final section concludes.

2.2 Model of Importing and Local Market Entry

In this section I lay out a model of retailer entry and look at the effect of imports on entry and exit decisions. There are two types of firms: single-unit and large firms. The firms differ in terms of their profit functions.

The model delivers two results. First, a fall in trade costs leads large firms to import more, expand into more markets, and earn higher profits in existing markets. Second, a fall in trade costs single unit firms earn lower profits and exit. The effect of these competing forces on competition is ambiguous and depends on the extent to which stores of large firms replace stores of small firms.

2.2.1 Model Setup

Firms play a static entry game in M markets. Markets differ in terms of a characteristic, $x_m \in X \subset \mathbb{R}$, such as population. A firm of type $t \in \{S, L\}$ must pay an entry cost $f_t > 0$ to enter a market. The profits of a firm of type t, $\pi^t(x, c, N_S, N_L, c_L, \epsilon)$, depend on the firm's marginal cost c, the number of firms of each type (N_S, N_L), the marginal cost of the large firm (c_L), the marginal cost of small firms (c_S), the firm's productivity (ϵ), and the productivity of all other firms $(\vec{\epsilon})$. I make the following assumptions about the profit functions

Assumption 1 *Each π^t satisfies the following conditions*

1. *π^t is increasing in x and ϵ*
2. *π^t is decreasing in c, N_S, N_L*
3. *π^S is increasing in c_L*
4. *π^t is continuous and differentiable in all variables*

The first assumption states that the characteristics of a market are defined such that they are desirable and that more productive firms are more profitable. The second assumption states that having higher costs or more competitors has a negative effect on profits. The third assumption states that small firms are substitutes with large firms. The fourth asusmption is made for convenience in proofs.

There are $\mathcal{E} = \{\mathcal{E}_S, \mathcal{E}_L\}$ firms of each type that could enter each market. I order firms of each type $s_1, s_2, \ldots, s_S$ and $l_1, l_2, \ldots, l_L$ by ϵ such that $\epsilon_s > \epsilon_{s+1}$ and $\epsilon_l > \epsilon_{l+1}$.

An equilibrium is a cutoff rule for $\hat{\epsilon}_{mS}$ and $\hat{\epsilon}_{mL}$ which implies a number of firms of each type $(\hat{N}_{mS}, \hat{N}_{mL})$ such that such that the following conditions are satisfied:

$$
\begin{aligned}
\pi^S(x_m, c, \hat{N}_{mS}, \hat{N}_{mL}, c_L, \epsilon_{\hat{N}_{mS}}) &\geq 0 \\
\pi^S(x_m, c, \hat{N}_{mS} + 1, \hat{N}_{mL}, c_L, \epsilon_{\hat{N}_{mS}+1}) &< 0 \\
\pi^L(x_m, c, \hat{N}_{mS}, \hat{N}_{mL}, c_L, \epsilon_{\hat{N}_{mL}}) &\geq 0 \\
\pi^L(x_m, c, \hat{N}_{mS}, \hat{N}_{mL} + 1, c_L, \epsilon_{\hat{N}_{mL}+1}) &< 0
\end{aligned}
$$

To avoid multiple equilibria I assume large firms enter first, followed by small firms. This timing assumption is for convenience. It ensures that when I compare equilibria before and after large firms start importing the differences are caused by changes in parameters instead of changes in the equilibrium I select. Given this assumption there is a unique equilibrium.

2.2.2 Testable Hypothesis

I consider comparative statics with respect to the marginal cost of large firms. An increase in direct importing lowers the marginal cost from c_L to $c_I < c_L$. The results are proven in the following proposition.

Proposition 1 *$\hat{\epsilon}_{mS}$ is non-decreasing in c_L and non-increasing in N_L.*

Proof Let $c_L < c'_L$ and $\hat{\epsilon}_{mS}$ be the cutoff rule with c_L then by assumption $\pi^S(x, c, N_s, N_L, c_L, \epsilon) < \pi^S(x, c, N_s, N_L, c'_L, \epsilon)$ for all ϵ. In particular for $\epsilon \geq \hat{\epsilon}_{mS}$ it must be that $0 \leq \pi^S(x, c, N_s, N_L, c_L, \epsilon) < \pi^S(x, c, N_s, N_L, c'_L, \epsilon)$. The proof for N_L is the same. ■

Proposition 2 *Let $\{\hat{\epsilon}_{mS}, \hat{\epsilon}_{mL}\}_{m=1}^{M}$ be an equilibrium where both types of firms have marginal cost c and $\{\bar{\epsilon}_{mS}, \bar{\epsilon}_{mL}\}_{m=1}^{M}$ be an equilibrium with importing then the following conditions must hold*

1. *$\bar{\epsilon}_{mL} \geq \hat{\epsilon}_{mL}$ for all m*
2. *$\bar{\epsilon}_{mS} \leq \hat{\epsilon}_{mS}$ for all m*

Proof By proposition 1 $\bar{\epsilon}_{mS} \leq \hat{\epsilon}_{mS}$ for all m which implies $\bar{N}_{mS} \leq \hat{N}_{mS}$. Then by assumption 1 for $\epsilon \geq \hat{\epsilon}_{mS}$

$$\pi^L(x, c_I, \bar{N}_{mS}, \hat{N}_{mL}, c_I, \epsilon) \geq > \pi^L(x, c_L, \hat{N}_{mS}, \hat{N}_{mL}, c_L, \epsilon) \geq 0$$

this implies that $\bar{\epsilon}_{mS} \leq \hat{\epsilon}_{mS}$ for all m. ■

These propositions deliver testable implications. They imply that a fall in trade costs should

1. Lead to the expansion of large firms.
2. Lead to the exit of stores of small firms.

These implications are tested in section 2.5.

2.3 Data: Retailer Revenue and Importing

This section describes the new data on store-level revenue in 20 product categories, which are combined with firm-level imports in these same product categories. These data allow for the construction of detailed measures of which stores compete with each other and the exposure of each store to direct imports.[3]

2.3.1 Data Description

This paper combines two primary sources of confidential U.S. Census Bureau microdata that cover 1992 to 2007, the time period during which imports of consumer goods grew substantially. The primary source of data is the Census of Retail Trade (CRT), which provides revenue by product type for retail establishments in years ending in 2 and 7. The CRT data are used to construct store-level revenue by 20 categories of goods called product categories. I use these data on revenue and information on the location of each store to define which stores compete with each other. Importantly, a store's local competition will include stores in many different industries. Store-level measures are combined with data on the activity of the firm that owns each store. In particular, information is tracked on how many stores each firm operates in each year and the trading activity of each firm.

The Longitudinal Foreign Trade and Transactions Database (LFTTD) is used for the collection of data on the imports of each retailer. These data contain the value of each firm's imports by source country and harmonized system (HS) code on a yearly basis. These data are used to construct firm-level imports in each product category.

The CRT and LFTTD are combined with the Longitudinal Business Data (LBD) to identify the activity of stores of each firm in other sectors of the economy. This information assists in the definition of a retail firm.

2.3.2 Sample Construction

A *retail firm* is defined as one that has at least 50 percent of its employment in retail and at least one store in the CRT.[4] The retail sector is defined based on the North American Industrial Classification System (NAICS) using the codes created by Fort and Klimek (2016). Three subsectors (3-digit NAICS codes) are removed: auto dealers and part stores

[3] I use store, shop, and establishment as synonyms.

[4] Almost all firms with any employment in retail have almost all of their employment in that sector. Unlike in manufacturing the biggest retailers are almost exclusively retail firms.

(441), gasoline stations (447), and non-store retailers (454).[5] The first two are removed because they have a large degree of franchising. Franchises are typically required to buy the products they sell at fixed prices from a parent company. Thus, any imports they sell come through a parent company and cannot be identified in the LFTTD.[6] Non-store retailers are removed because they reach consumers primarily through the internet and catalogs. This prevents identifying in which markets they sell their products. The non-store retail sector accounted for less than 10 percent of sales in all eight of the major product categories prior to 2007, the final year of this study. This leaves eight subsectors: furniture and home furnishings stores (442); electronics and appliances stores (443); home goods and gardening stores (444); food and grocery stores (445); health goods stores (446); clothing and apparel stores (448); toy, hobby, and sporting goods stores (451); general merchandise stores (452); and miscellaneous store retailers (453).

Retail firms are partitioned into three types: large firms, small chains, and single units. Large firms are defined as firms with more than 100 stores in the retail sector. These firms typically operate in many markets and have many more than 100 stores.[7] Small chains have between 2 and 99 stores. The majority of stores of small chains belong to firms with fewer than 10 stores. Single unit firms have only one store in the retail sector. Small chains and single unit firms are collectively refferred to here as *small firms.* I will refer to the type of stores based on the type of the firm to which they belong.

2.3.3 Creation of Revenue by Product Category

The CRT asks establishments to provide data on revenue by product line (for example, men's footwear, women's pants, diamond jewelry). Unlike in other sectors of the economy, retail stores compete with stores in other industries. In particular, general merchandise stores such as Walmart and Target, compete with stores in groceries and electronics. Thus, revenue by product line is important when looking at competition in the retail sector. The product line codes are aggregated into 20 product categories such that stores in industries outside of general merchandise sell primarily in one product category. For instance, stores in subsector 448 (clothing and clothing accessory stores) primarily report sales in products such as women's dress pants, men's suits, and footwear, which are grouped into a clothing product category. Table B.1.2 lists these product categories.

[5] The subsectors in my sample account for 58 percent of total NAICS-based retail sales in 2007 or $2.2 trillion in sales.

[6] Additionally, forces that benefit large firms, such as improvements in communication technology are unlikely to help franchises in the same way, making them unsuitable as a control group.

[7] Table 2.1 shows that the average large firm has over 600 stores in 2007. For comparison Walmart reported it operated about 4,000 stores in its 2007 Annual Report.

Aggregating data in this allows for accurately imputing revenue by product category for stores that do not report product line data. The CRT only asks for detailed product lines from a sample of small stores. For the remainder, store-level revenue estimates are constructed from administrative data, without revenue by product line. This affects stores accounting for 20 percent of sales. For these stores, the distribution of their sales across product categories are imputed using characteristics of the store, such as industry and multi-unit status. Details of this procedure are provided in Appendix B.1.1.

2.3.4 Creation of Imports by Product Category

The product category revenue data are then combined with imports by product category from the LFTTD. The HS codes provided in the data are mapped to product categories. This is accomplished by updating the concordance from Basker and Van (2010) using data on which products retailers import. The concordance is constructed such that almost all products retailers import map to a product category. The HS codes that can be mapped to a product category are called consumer goods HS codes.[8] Details are provided in Appendix B.1.2.

Imports by product category are matched with firm-level sales by product category to get import penetration by product category for each firm. It is important to define import penetration by product category because of the existence of significant variation in import penetration across product categories within a firm. For example, a firm may import almost all of its clothing directly, but not import any groceries. This variation makes it feasible to separately control for competition with large firms and competition with direct imports.

2.4 Increasing Importance of Large Retailers

Large firms are increasingly dominating the retail sector. This section shows some statistics quantifying the size of these changes with a specific focus on the entry and exit of stores of different size firms. Then it describes the importing behavior of retailers arguing that

2.4.1 Expansion and Exit

The difference between national and local concentration is driven by large firms' extensive margin decisions to enter new markets. Table 2.1 shows that between 1997 and 2007 the

[8]Proceeding in this way causes me to overstate consumer goods imports by non-retailers. For example, if some retailers import desks, I classify all desks as consumer goods imports, though some desks are imported by wholesalers for sale to businesses or government entities. This only affects statistics on the fraction of consumer goods imports by retailers.

number of stores of large retailers increased significantly as large retailers expanded into new markets while many small stores exited.

Table 2.1: Number of Stores by Firm Type

	Stores		
Firm Type	1997	2007	Percent Change
Single Unit	409,655	392,027	-4
Small Chains	107,811	89,042	-17
Large Firms	177,139	220,222	24
Total	694,605	701,291	1
	Firms		
Single-Unit	409,655	392,027	-4
Small Chains	21,432	17,209	-17
Large Firms	368	349	-5
Total	431,455	409,585	-5

Notes: The figures come from author calculations using the public CRT. They represent the sum of subsectors 442, 443, 444, 445, 446, 448, 451, 452, and 453, and include single-unit firms (with one store), small chains (with 2 - 99 stores), and large firms (with 100+ stores).

The fact that the number of stores by large firms increased by 24 percent, but the number of large firms was essentially unchanged, decreasing from 368 to 349, implies the increase in stores of large firms was caused by an increasing number of stores per large firm. Furthermore, these new stores were built in new markets. In 1997, the average large firm had stores in 114 commuting zone. This number increased by 26 percent to 145 by 2007. As these large firms expanded, smaller stores exited. In particular, the number of small chains decreased by 17 percent. On average, single-unit stores also exited, but the decrease was much smaller. The number of single-unit stores decreased by four percent.

2.5 Direct Imports and the Retail Sector

Evidence is now presented concerning a link between direct importing and increasing concentration. First, the increase in imports between 1997 and 2007 is quantified. Second, changes in the share of large firms are related to direct imports, with mixed results. Finally, it is demonstrated that small stores exposed to competition from direct imports are more likely to exit.

This subsection establishes a few relevant patterns in the data. First, I establish the importing directly is a behavior that is undertaken primarily by the largest retail firms, with the smaller firms that do import importing only a small fraction of their sales. This is used as justification for the assumption that small firms cannot import. Second, I show the importance of China in retailer imports and argue that focusing on imports from China is good.

This paper focuses on imports of consumer goods from China for two reasons. First, China has been the main source of retailer imports for the last 20 years. Second, most of the growth in consumer goods imports in the 2000s was accounted for by increasing imports from China. Table 2.2 shows the top sources of retailer imports ranked by value and by number of retail firms that import from a country. China is consistently ranked number one by value of imported products, although Canada have more firms importing from it.

Table 2.2: Countries by Rank

	Rank by:	
	Value	Firm
China	1	2
Italy	2	3
Canada	3	1
India	4	6
Indonesia	5	17

Notes:

The second fact that I exploit is that only the largest retailers.

2.5.1 Direct Imports and Expansion of Large Firms

Imports by retailers increased drastically between 1997 and 2007, as did imports by non-retailers of consumer goods. Table 2.3 shows the change in imports of consumer goods from all countries and from China in particular. Total imports of consumer goods increased from 538 billion to 1.2 trillion U.S. dollars between 1997 and 2007. Imports from China accounted for 40 percent of this increase. Twenty percent of imports from China were by retailers and over 90 percent of those imports were by large firms.

Large retailers directly import a significant fractions of their sales. Table 2.4 shows imports of large firms divided by sales of large firms by department. Between 1997 and 2007 large firms significantly increased the fraction of sales imported directly in departments such as

Table 2.3: Imports of Consumer Goods

	1992	1997	2002	2007
Total Consumer Goods Imports	319	538	794	1,192
Imports from China	23	55	124	316
Retailer Imports from China	6	13	27	62
Large Retailer Imports from China	5	12	25	57

Notes: The data are from the LFTTD micro data set. Consumer goods are defined as HS codes that map to a retail department using the procedure described in Appendix B.1.3. The values are in billions of 2007 U.S. dollars.

furniture, toys, electronics and appliances, and clothing. Imports are particularly important in departments such as clothing and clothing accessories, where by 2007 they represented 17 percent of all sales. This translates to more than 35 percent of cost of goods sold.[9] This is in share contrast to the grocery store subsector, where imports are less than one percent of sales.

Table 2.4: Share of Large Firm Sales Imported Directly

	1997	2007
Clothing	10.3	16.2
Electronics and Appliances	5.0	12.5
Furniture	11.6	47.8
Groceries	0.1	0.4
Health Goods	0.4	0.6
Home Goods	2.7	7.4
Sporting Goods	7.3	10.0
Toys	16.4	30.1

Notes: The figures represent total imports by large firms in a department divided by total sales by large firms in that department multiplied by 100.

A key challenge in relating the expansion of large retailers to importing is that large retailers often sell many products. For example, during the 1990s and 2000s Walmart opened thousands of supercenters selling groceries, clothing, and other goods. The entirety of this

[9] Cost of goods sold is typically 60 percent of sales, but includes spending on domestic transportation and other domestic costs of foreign goods. Thus, the sales share of imported goods should be at least twice the ratio of value of imports to sales.

expansion was surely not due to direct imports—Walmart's expansion started well before the increase in direct imports.[10] Additionally, other large retailers such as CVS and Walgreens added thousands of stores despite a relatively low degree of direct importing.

I investigated whether large firms that sold products exposed to direct imports expanded between 1997 and 2007 and found mixed results. First, I compared the change in the share of sales going to large firms to the share of sales imported directly. The results are plotted in Figure 2.1. Six departments followed a clear pattern: that larger increases in the share of sales imported directly are related to increasing share of sales by large firms. However, there are two major exceptions, toys and furniture. Toys began as the most concentrated department with 84 percent of sales by large firms so even if imports had a large effect there was very little room for growth. In contrast, furniture is the department with the lowest share of sales by large firms. It is also the department where large firms account for the smallest share of total imports. Many small furniture firms import. This may imply large retailers have less of a marginal cost advantage in furniture. Thus, the evidence of a relationship between changes in direct import share and changes in the share of large firms depends on the product in question.

2.5.2 Direct Imports and Exit of Small Stores

There has been considerable research undertaken linking the exit of small stores to competition with Big-Box stores and Walmart in particular.[11] A main contribution of this current research is to assess the role of direct imports as a channel through which large firms cause the exit of small stores. To do so, the relationship between the exit of small stores and increasing exposure to direct imports is estimated using data from 2002 and 2007.

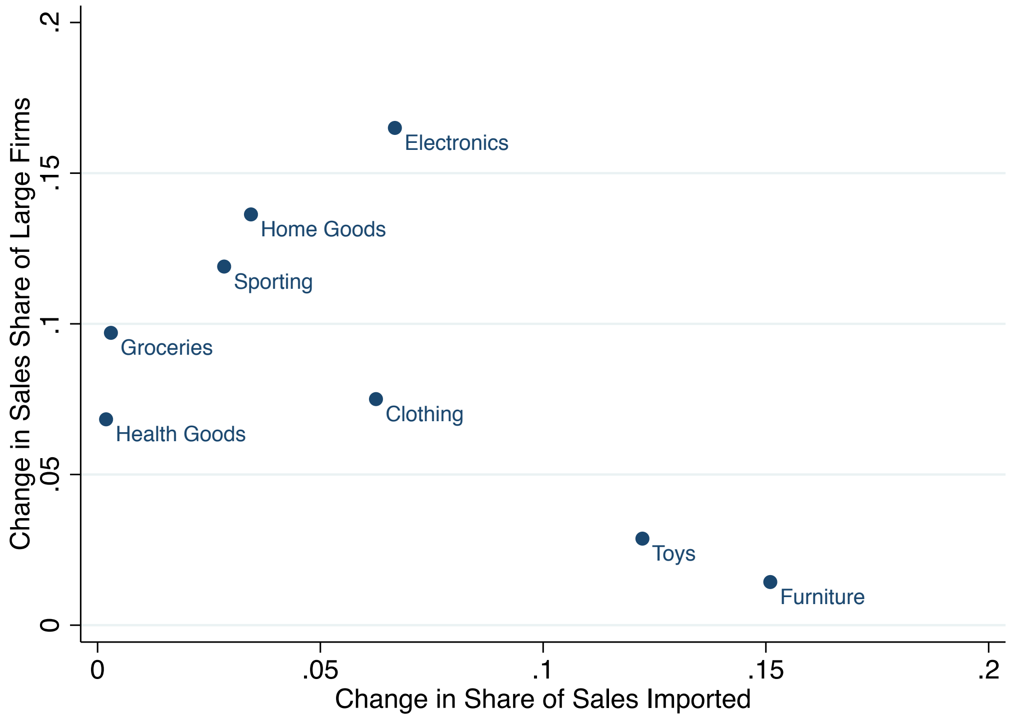
To estimate the relationship between competition with direct imports and exit of small stores a store-level measure of exposure to direct imports is developed. For each small store, I calculate the fraction of competitors' sales that are imported directly. Data on both location and sales by department are used to define competition between stores.

I estimate the effect of changes in import exposure on the probability of exit of small stores. Regressions are run separately for single-unit stores and small chains because these two types of stores may face different degrees of competition from imports. In particular, single-unit stores may sell niche products that are less subject to competition from imports. The estimating equation is:

[10] See Basker (2005) or Holmes (2011).

[11] For example, see Basker (2005); Haltiwanger et al. (2010); Jia (2008); Arcidiacono et al. (2016). I confirm that in my sample I also find that small stores competing with large retailers are more likely to exit.

Figure 2.1: Share of Sales by Large Firms vs Direct Import Share - Department-Level



Notes: The data are from the CRT and LFTTD micro data sets. The fraction imported is the value of direct imports divided by the value of sales. Sales share of large firms is the sales of firms with more than 100 stores divided by total sales in a department. Changes are calculated between 1997 and 2007.

$$E_{im}^{2002-2007} = \alpha + \beta_1 \Delta d_{im}^{2002-2007} + \beta_2 d_{im}^{2002} + X_{im}\Gamma' + \varepsilon_{im}. \tag{2.1}$$

$E_{im}^{2002-2007}$ is an indicator of whether store i in market m exits between 2002 and 2007. $\Delta d_{im}^{2002-2007}$ is the change in exposure to imports for that store and d_{im}^{2002} is the 2002 level of exposure to direct imports. X contains a number of controls for store characteristics, market characteristics, and the competitive environment of each store.

The controls for store characteristics include dummy variables for store age and the log of the store's 2002 sales. I also include dummy variables for each store's top department, the one in which the store has the plurality of its sales. This controls for national demand shocks to certain departments and differences in the natural rate of turnover of stores based on what they sell. The controls for market characteristics include the population of the market and the change in population of the market between 2002 and 2007. Finally, X contains controls for the size and scope of the competitors of each store in 2002. The construction of those controls is described in subsection 2.5.2.

Creation of Direct Import Exposure

The creation of direct import exposure, $\Delta d_{im}^{2002-2007}$, takes place in three steps. The first step involves calculating *firm-department-level* import penetration at the national level for all retailers using the imports of each firm. The second step involves using the distribution of stores of each firm to create *department-market-level* import penetration. The final step is to create *store-level* import exposure by weighting department-market-level import penetration using each store's sales in each department.

First, I calculate firm-department-level import penetration for all retailers. For firm, k, the fraction of sales in each department, j, in year, t, that are imported directly is

$$dimpen_{kjt} = \frac{imports_{kjt}}{sales_{kjt}}. \tag{2.2}$$

I assume that each firm's imports in a department are distributed across its stores according to that store's share of firm sales in each department. That is, I assume if Walmart imports 10 percent of its clothing directly, 10 percent of the clothing sales of each of its stores are imported directly.

In the second step, I calculate a measure of direct import penetration in department j in market m at time t as

$$dimpen_{mjt} = \sum_{k=1}^{K} s_k^{jmt} dimpen_{kjt}, \tag{2.3}$$

where s_k^{jmt} is the sales share of firm k in department j in market m in year t.

This market-department-level measure of direct import competition is converted to the store level by weighting each market-department-level direct import penetration according to the sales share of that department in the store's total sales, s_j^{imt}. $k(i)$ is defined to be the firm of store i. I remove the contribution of that firm from market-level import penetration and rescale the shares of the other firms so that direct import penetration is the fraction of each store's competitors' sales that are imported directly. The resulting measure of direct import exposure for store i at time t is:

$$d_{im}^{t} = \sum_{j=1}^{J} s_j^{imt} dimpen_{jmt}^{-k(i)}. \tag{2.4}$$

It is useful to understand how changes in the import exposure measure relate to outcomes of small stores such as their probability of exit and sales growth between 2002 and 2007. So I calculate the change in import exposure, $\Delta d_{im}^{2002-2007}$.

Table 2.5 shows summary statistics for both samples in the regression. The average single-unit store experienced an increase in import exposure of 0.01 from an initial level of 0.012. The average small chain experienced an increase in import exposure of 0.012 from an initial level of 0.014. This number may seem small, but it is depressed significantly by a large number of grocery stores and pharmacies that experienced almost no competition from direct imports. Thus, the change in exposure to direct imports for clothing, furniture, and electronics stores is significantly higher.

Table 2.5: Summary Statistics - Exit and Exposure of Small Stores

	Single-Unit		Small-Chain	
	Mean	S.D.	Mean	S.D.
Change in import exposure ($\Delta X_{im}^{2002-2007}$)	0.010	0.015	0.012	0.017
Import exposure (X_{im}^{2002})	0.012	0.013	0.014	0.014
Exposure to large firms (pct_{im}^{L})	0.518	0.176	0.535	0.188
Exposure to GMs (pct_{im}^{GM})	0.238	0.140	0.249	0.139
Exposure to small chains (pct_{im}^{SC})	0.202	0.189	0.192	0.118
Probability of exit ($E_{im}^{2002-2007}$)	0.469	0.499	0.356	0.479
Number of observations	488,000		87,000	

Notes: Summary statistics are for the sample of single-units and small chains. The observation count was rounded to the nearest thousand. Unless otherwise indicated, variables are calculated for the year 2002. Import exposure is the fraction of competitors' sales that are imported directly. Exposure to types of firms is the percentage of competitors' sales that are by a firm of a given type as described in subsection 2.5.2. Exposure to GMs is the fraction of competitors' sales by general merchandise firms. Probability of exit is the probability a store closes between 2002 and 2007.

I expect increases in competition from direct imports to lead to the exit of small stores. However, an OLS regression of exit on the change in direct import exposure will be biased if competing firms' decisions to directly import are correlated with other activities that lower costs or improve quality. It will also be biased if importers enter markets with stores that are likely to exit.

Instrumental Variables Strategy

These concerns about bias are addressed using an instrumental variables (IV) strategy. The IV strategy identifies firms that increased imports because of China's increase in exports to other countries.[12] Specifically, I used each firm's 2002 imports by 6-digit HS code combined with the growth rate of exports from China in that same product code to construct the change in exposure due to increased exports from China.

In 2002, some retailers in the sample were already importing certain products from China. Between 2002 and 2007, China's exports of some of these products grew substantially while exports of other products did not. This growth had the effect of increasing imports by some retailers more than others. I exploit the variation in the change in direct import exposure faced by each store due to growth in China's exports. The instrument relies on the assumption that retailers importing products in 2002 that subsequently grew were not also becoming increasingly competitive for other reasons.

I describe the instrument in two steps. In the first step, I calculate firm-department-level import penetration in 2007 using the growth of exports from China. In the second step, the firm-department-level import penetration is converted to a store-level measure of import exposure using sales share information from 2002.

The measure of import penetration of each firm in a department in 2007 uses the firm's imports by six-digit HS code in 2002, the growth rate of Chinese exports to high-income countries between 2002 and 2007, and the U.S. growth rate of department-level sales between 2002 and 2007. It is calculated as

$$\hat{dimpen}_{kj2007} = \frac{\sum_{h=1}^{H_J} imports_{kh2002}(1 + g_{h,CN\rightarrow HI}^{2002-2007})}{sales_{kj2002}(1 + g_{j,US}^{2002-2007})}. \quad (2.5)$$

h is an individual 6-digit HS code and $g_{h,CN\rightarrow HI}^{2002-2007}$ is the growth rate of Chinese exports of product h to other high-income countries between 2002 and 2007. H_j is the set of HS codes that are matched to department j using the procedure in Appendix B.1.3. $g_{j,US}^{2002-2007}$ is the growth rate of all U.S. sales in department j. Thus, the numerator of equation (2.5) is firm imports in department j in 2007 if imports of each product grew at the rate of China's exports. $\hat{dimpen}_{kj2007}$ is what each firm's import penetration in a department would have been in 2007 had its imports in each product grown at the rate of China's exports to other

[12]The other countries are Australia, Denmark, Finland, Germany, Japan, New Zealand, Spain, and Switzerland.

countries and its sales had grown at the rate of national sales.

These measures are combined to calculate each store's competition from direct imports as the weighted average of predicted firm-department-level direct import penetration, holding each store's competitor's shares at the 2002 level. Doing so eliminates the effect of entry of additional importers from the instrument, meaning the instrument only captures intensive margin changes. The measure is defined as

$$Z_{im}^{2007} = \sum_{j=1}^{J} s_j^{im2002} \sum_{k=1}^{K} s_k^{jm2002} dimpen_k^{j2007}. \tag{2.6}$$

The change in $Z_i m$ is defined as

$$\Delta Z_{im}^{2002-2007} = Z_{im}^{2007} - X_{im}^{2002}$$

In the first stage of IV, I estimate the relationship between the change in actual store-level direct import exposure and the change in exposure due to growth in exports from China. The specification is

$$\Delta d_{im}^{2002-2007} = \alpha + \beta_t \Delta Z_{im}^{2002-2007} + \beta_1 d_{im}^{2002} + X_{im}\Gamma_1' + \varepsilon_{im}, \tag{2.7}$$

where $\Delta Z_{im}^{2002-2007}$ is the change in import exposure using growth of exports from China and d_{im}^{2002} is the exposure of store i in 2002. The regression contains the same controls as in equation (2.1).

Controls for Competitors' Size and Scope

A significant concern with this regression is that exposure to direct imports is correlated with exposure to large firms. In particular, the locations that are predicted to have larger increases in exposure between 2002 and 2007 may be the locations with more large firms in 2002. Controls for competition with small chains and large firms are included for this reason.

For each store, I define the fraction of sales by competitors that are by small chains (SC) and large firms (L). For competitor type $w \in \{SC, L\}$ the fraction of sales by competitors that belong to firm type w are

$$pct_{im}^{w} = \sum_{j=1}^{J} s_j^{im2002} s_w^{jm2002,-k(i)}, \tag{2.8}$$

where $s_w^{jm2002,-k(i)}$ is the share of firms of type w in department j and market m in 2002 with the sales of the firm of i removed. Thus, pct_{im}^{w} takes the share of competitor's sales by firms of type w in department j and weights it by the share of department j in store i's sales. In these controls, the left out group is single-unit stores. Thus, the coefficients pct_{im}^{L} and pct_{im}^{GM} represent the impact of competing with a large firm or small chain instead of a single-unit store.

The control for large firms includes both large firms that primarily sell one type of product and general merchandisers. I include an additional control for general merchandisers because general merchandisers may sell products that are more or less substitutable with those sold by small stores. They also may be more likely to import. Exposure to general merchandisers is defined as

$$pct_{im}^{GM} = \sum_{j=1}^{J} s_j^{im2002} s_{GM}^{jm2002,-k(i)}. \tag{2.9}$$

Large general merchandisers account for almost 100 percent of sales in the general merchandising industry by 2002 so the coefficient on pct_{im}^{GM} essentially measures the difference in exit probability between competing with a large firm outside of general merchandising and competing with a large general merchandiser.

I find that most stores already had fairly high exposure to large firms and that increased significantly over the five years. Table 2.5 shows that in 2002, 52 percent of sales by the competitors of the average single-unit were by large firms. The number is similar for small chains. Roughly half of the exposure to large firms comes from general merchandisers. Specifically, 23.8 percent of the sales of competitors to single-units and 24.9 percent of the sales of competitors to small chains were by general merchandisers.

OLS and IV Results

Table 2.6 shows that as expected the change in import exposure due to growth in exports from China is a significant predictor of the actual change in import exposure. Column 1 implies that a one percentage point increase in predicted import exposure is associated with a 0.18 percentage point increase in actual exposure.

Table 2.6: First Stage Results

Dependent Variable is the Change in Direct Import Exposure ($\Delta d_{im}^{2002-2007}$)		
	Single-Unit	Small Chain
$\Delta Z_{im}^{2002-2007}$	0.175***	0.173***
	(0.008)	(0.014)
d_{im}^{2002}	-0.450***	-0.091*
	(0.037)	(0.047)
pct_{im}^{L}	0.010***	0.008**
	(0.001)	(0.002)
pct_{im}^{GM}	-0.015***	-0.017***
	(0.002)	(0.002)
pct_{im}^{SC}	0.015***	0.013***
	(0.002)	(0.002)
Log Sales	0.000	0.000
	(0.000)	(0.000)
Top Department Fixed Effects	Y	Y
Age Fixed Effects	Y	Y
Market Controls	Y	Y
R2	0.64	0.66
Observations	488,000	87,000

Notes: $\Delta Z_{im}^{2002-2007}$ is the change in import exposure of store i using exports from China. $\Delta d_{im}^{2002-2007}$ is the change in direct import exposure of the store between 2002 and 2007. d_{im}^{2002} is the level of exposure in 2002. pct_{im}^{w} is the exposure of store i to firms of type $w \in \{$Large, General Merchandiser, Small-Chain$\}$. Regressions include fixed effects for top department of each store and for the age of the store. Market controls are the commuting zone population in 2002 and the change in commuting zone population between 2002 and 2007. Standard errors are clustered at the commuting zone, top department level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The observation count is rounded to the nearest thousand.

Table 2.7 shows results for the main specification for both OLS and IV with the regression run separately for both types of stores. In the single-unit sample, Positive and significant coefficients in both OLS and IV were found. The IV results suggest a one percentage point increase in exposure leads to a 0.78 percentage point increase in the probability of exit of small stores. Table 2.5 showed 47 percent of single-unit stores exited during this period and the average change in exposure was one percentage point. Combining these facts indicate increasing exposure to direct imports explains a small fraction of the exit of single-unit stores.

In Table 2.7, column 4 shows that competition with direct imports has a larger effect on small chains. A one percentage point increase in exposure to direct imports increases the exit probability of small chains by 1.7 percentage points. This effect is significantly larger given that the exit probability of small chains was 36 percent during this period.This result implies direct imports increased the exit probability of small chains by about 5 percent. Additionally, I find much larger coefficients on the level of exposure each store faced in 2002 for small chains than in single-unit stores.

The IV results are likely an understatement of the impact of direct imports on the exit probability of small stores because the instrument holds the entry decisions of large stores fixed. If importing directly causes large stores to expand then that effect is not captured by the instrument.

Additional Results and Controls

Here I consider specifications with the growth rate of sales as the dependent variable and find that small stores that did not exit had slightly higher growth indicating they may have benefited from decreased competition from other small stores. This motivates the use of the model in the next section to consider the net effect of direct imports.

The first additional specifications I consider have different dependent variables. In particular, I consider the percent change in sales between 2002 and 2007 and the change in sales measured by the Davis-Haltiwanger-Schuh (DHS) growth rate (Davis, Haltiwanger and Schuh, 1996) which is defined as

$$y_{im}^{DHS,2002-2007} = 2\left(\frac{sales_{im}^{2007} - sales_{im}^{2002}}{sales_{im}^{2007} + sales_{im}^{2002}}\right).$$

This measure is the change in sales divided by average sales over the two years. It takes values between -2 and 2 and can be calculated for stores that exit as well as stores that do not. In this way, I can test whether continuing stores also grow less. Table 2.8 shows that between 2002 and 2007 real sales of both types of small stores decreased. Columns 1 and 3 show the average DHS growth rate was -0.96 for single-units and -0.73 for small chains. Columns 2 and 4 show real sales of single-units and small chains decreased by 5.8 percent and 4.3 percent, respectively, over the five-year period.

Table 2.9 shows results with this measure. I find that increases in import exposure lower the DHS growth rate, but I do not find lower sales growth rates for stores that survived. This may be due to the fact that the small stores that survived faced less competition from other small stores. Another contributing factor might be that the stores that survived may have been more horizontally differentiated. Appendix B.1.5 presents additional specifications, including: results with controls for whether a small store imports and results with the change in exposure to large firms and general merchandisers.

Table 2.7: Exit of Small Stores

Dependent Variable is an Indicator of Whether a Store Exits Between 2002 and 2007				
	Single-Unit		Small Chain	
	OLS	IV	OLS	IV
$\Delta d_{im}^{2002-2007}$	1.006***	0.775**	1.006***	1.728**
	(0.129)	(0.325)	(0.249)	(0.805)
d_{im}^{2002}	0.255	0.488**	0.989**	1.224***
	(0.181)	(0.232)	(0.451)	(0.464)
pct_{im}^{L}	0.066***	0.125***	0.042	0.105***
	(0.019)	(0.011)	(0.029)	(0.027)
pct_{im}^{GM}	0.011	-0.109***	-0.056	-0.163***
	(0.020)	(0.018)	(0.038)	(0.038)
pct_{im}^{SC}	0.068***	0.104***	0.004	0.014
	(0.019)	(0.017)	(0.035)	0.037)
Log Sales	-0.101***	-0.101***	-0.082***	-0.081***
	(0.001)	(0.001)	(0.002)	(0.002)
Top Department Fixed Effects	Y	Y	Y	Y
Age Fixed Effects	Y	Y	Y	Y
Market Controls	Y	Y	Y	Y
R2	0.122	0.121	0.065	0.064
Observations	488,000	488,000	87,000	87,000

Notes: Dependent variable is $E_{im}^{2002-2007}$. Instrument is, $\Delta Z_{im}^{2002-2007}$, the change import exposure of store i using exports from China. $\Delta d_{im}^{2002-2007}$ is the change in direct import exposure of the store between 2002 and 2007. d_{im}^{2002} is the level of exposure in 2002. pct_{im}^{w} is the exposure of store i to firms of type $w \in$ {Large, General Merchandiser, Small-Chain}. Regressions include fixed effects for top department of each store and for the age of the store. Market controls are the commuting zone population in 2002 and the change in commuting zone population between 2002 and 2007. Standard errors are clustered by commuting zone and top department. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The observation count is rounded to the nearest thousand.

Table 2.8: Summary Statistics - Growth of Small Stores

	Single-Unit		Small-Chain	
	All	Continuers	All	Continuers
Mean	-0.960	-0.058	-0.730	-0.043
Standard Deviation	1.070	0.822	1.037	0.683
Observations	488,000	259,000	87,000	56,000

Notes: Summary statistics are for the sample of single-units and small chains. The observation count is rounded to the nearest thousand. "All" is computed using the D.H.S growth rate between 2002 and 2007 calculated according to Davis et al. (1996). "Continuers" is calculated using the percent growth rate.

Table 2.9: Growth of Small Stores

Dependent Variable is the Growth Rate of Sales				
	Single-Unit		Small Chain	
	All	Continuers	All	Continuers
$\Delta d_{im}^{2002-2007}$	-1.268*	1.607*	-5.257***	-2.491
	(0.688)	(0.861)	(1.813)	(-1.965)
d_{im}^{2002}	-0.752	1.236***	-1.878*	0.803
	(0.472)	(0.457)	(0.965)	(1.418)
pct_{im}^{L}	-0.328***	-0.183***	-0.302***	-0.181***
	(0.024)	(0.023)	(0.061)	(0.047)
pct_{im}^{GM}	0.166***	-0.135***	0.305***	0.007
	(0.037)	(0.030)	(0.082)	(0.074)
pct_{im}^{SC}	-0.281***	-0.167***	0.009	0.052
	(0.035)	(0.032)	(0.080)	(0.061)
Log Sales	0.171***	-0.110***	0.129***	-0.072***
	(0.002)	(0.003)	(0.005)	(0.006)
Top Department Fixed Effects	Y	Y	Y	Y
Age Fixed Effects	Y	Y	Y	Y
Market Controls	Y	Y	Y	Y
R2	0.073	0.094	0.043	0.049
Observations	488,000	259,000	87,000	56,000

Notes: The "All" columns contain all stores with sales growth calculated using the DHS growth rate between 2002 and 2007 calculated according to Davis et al. (1996). The "Continuers" columns contain only stores that existed in both 2002 and 2007, with sales growth calculated using the log difference in sales between 2002 and 2007. $\Delta Z_{im}^{2002-2007}$ is the predicted change import exposure of store i using exports from China. $\Delta d_{im}^{2002-2007}$ is the change in direct import exposure of the store between 2002 and 2007. d_{im}^{2002} is the level of exposure in 2002. pct_{im}^{w} is the exposure of store i to firms of type $w \in$ {Large, General Merchandiser, Small-Chain}. Market controls are the commuting zone population in 2002 and the change in commuting zone population between 2002 and 2007. Standard errors are clustered at the commuting zone-top department-level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The observation count is rounded to the nearest thousand.

2.6 Conclusion

Chapter 3

Foreign Sourcing and Local Retail Concentration

3.1 Introduction

In the past 30 years, U.S. retailing has become substantially more concentrated at both the national and local level (Smith and Ocampo, 2019). A major contributor to this change has been the replacement of small retail firms with stores of large retail firms. The exit of these small retail firms has been linked to increasing imports of consumer goods from China by Smith (2019). These findings have the potential to miss important aspects of the effect of direct imports. In particular, they do not account for the endogenous response of entry rates of stores of all sizes to changes in trade.

This paper uses a dynamic continuous-time model of store entry and exit following Arcidiacono, Bayer, Blevins and Ellickson (2016) to partially account for these factors. Crucially, the model allows for a market-level import state. The evolution of this state flexibly depends on the size of the stores in the market. The model is estimated for clothing stores, an industry where direct imports are particularly important, with results indicating that direct imports have a negative impact on the profits of small stores, but do not lead to increasing concentration.

These estimates are used to simulate the structure of the retail sector without the observed increase in direct imports. The counterfactual accounts for two ways in which direct imports changed the structure of the retail sector. First, it accounts for the exit of small stores due to direct competition with products that were imported directly. Second, it partially accounts for increased entry rates of large stores due to increased profits conditional on entering. I find a relatively small importance of the direct channel. Removing only the first channel increases the number of small stores by six percent. The number of small clothing stores decreased by 30 percent during this period, which implies trade accounts for at least 14 percent of the decrease. However, no change in local concentration is found because large stores replace the exiting small stores.

The key challenge in this paper is simultaneously modeling importing decisions and local retail markets. Retail firms make importing decisions at the firm level based on the competitive environment they face in each of the markets in which they have a store. Modeling both choices simultaneously would eliminate the ability to treat each market independently of all others which is the typical assumption in the dynamic entry game literature (Aguirregabiria and Nevo, 2011). I maintain the independence assumption by modeling imports as a market-level state that evolves according to the number and type of stores in a market. This assumption maintains tractability while accounting for the level of import competition that each store faces. I show that under mild conditions these estimates are a lower bound on the effect of importing.

Another innovation of this paper is focusing on the intensive margin of importing. I measure import exposure as the fraction of competitor's sales that are imported directly, instead of the fraction of competitors that import anything. These two measures differ substantially because many retailers import a small amount of their products directly while a few large retailers import a significant fraction of their sales directly. This paper focuses on the fraction of competitors' sales that is directly imported because this fraction should correspond to the amount competitors' save by avoiding intermediary markups.

The results indicate that increasing direct imports played a substantial role in causing the exit of small stores. For single-unit stores, a one percentage point increase in the fraction of goods imported directly in a market increases the probability of exit of a store that was previously indifferent between entering and exiting by 14 percentage points. This is a significant effect given that nationally the fraction of clothing sales that were imported directly increased from three to six between 1997 and 2007. I find these effects are also large relative to the effect of entry by an additional general merchandise store, which increases the probability of exit by an indifferent store by six percentage points.

This study contributes to the literature on concentration in the retail sector by estimating the contribution of importing by retailers in affecting local concentration. A growing body of work has shown increasing national concentration and the declining labor share (Autor et al., 2017), as well as declining churn and reallocation of aggregate activity to large established firms (Decker et al., 2014). Despite a broad consensus on increased concentration, there is little evidence on the mechanisms driving the change. This paper contributes by focusing on a specific sector, retail, in which the growth in aggregate concentration has been particularly dramatic.[1] I show it is important to distinguish between local and aggregate concentration when thinking about these trends. In particular, an investigation is undertaken on the role of globalization in benefiting large retailers by providing them with direct access to cheap goods. The evidence does not suggest that imported inputs are a driving force behind local concentration in the retail sector.

I contribute to ongoing research on the role of foreign sourcing by studying how it impacts the decisions of retail firms. Previous work has shown that retailers play an important role in imports from China (Bernard, Jensen, Redding and Schott, 2010). Ganapati (2018) shows that intermediary markups can be substantial, which implies the marginal cost savings of importing directly can be large. The focus of this paper is on the domestic impact of direct imports by large firms on small retailers. My results complement work by Meinen and Raff (2018) who find that increases in aggregate industry-level direct imports are associated with the exit of small stores. This study's dataset allows for the construction of a store-level measure of exposure to direct imports. The results are consistent with Meinen and Raff (2018). Conditions are established under which the results are a lower bound on the effect of importing. The tightness of this bound depends on the magnitude of complementaries across activities of large retailers, such as importing, exporting, and entry as documented in Antras, Fort and Tintelnot (2017) and Bernard, Jensen, Redding and Schott (2018).

The rest of the paper proceeds as follows. Section 3.2 describes the data. Section 3.3 lays out a model of retailer competition. Section 3.4 discusses the estimates of the profit function of small stores. Section 3.5 estimates of the effect of removing direct imports on the structure of retail. The final section concludes.

3.2 Data: Retailer Revenue and Importing

This paper combines two primary sources of confidential U.S. Census Bureau microdata that cover 1997 to 2007, the time period during which imports of consumer goods grew substantially. The primary source of data is the Longitudinal Business Database (LBD), which contains employment and industry data for every store in the U.S. on a yearly basis. Unfortunately, the data do not contain sales so I define clothing stores by the industry reported by that store. Almost all clothing sales are accounted for by either general merchandisers or stores in the clothing subsector (NAICS 448). Further details on sample construction are detailed in Appendix B.1.4.

[1] Autor et al. (2017) find the national HHI in retail doubles between 1997 and 2007.

The Longitudinal Foreign Trade and Transactions Database (LFTTD) is used for the collection of data on the imports of each retailer. These data contain the value of each firm's imports by source country and harmonized system (HS) code on a yearly basis.

3.2.1 Creation of Department-Level Imports

Imports of clothing are identified using the concordance from Basker and Van (2010) updated to include most products imported by clothing retailers.

3.2.2 Sample Construction

The retail sector is defined based on the North American Industrial Classification System (NAICS) using the codes created by Fort and Klimek (2016).

The firms to which retail establishments belong are partitioned into three types: large firms, small chains, and single units. Large firms are defined as firms with more than 100 stores in the retail sector. These firms typically operate in many markets and have many more than 100 stores. Small chains have between 2 and 99 stores. The majority of stores of small chains belong to firms with fewer than 10 stores. Single unit firms have only one store in the retail sector. Small chains and single unit firms are collectively referred to here as *small firms*. I will refer to the type of stores based on the type of the firm to which they belong. For computational reasons, the focus is on 219 commuting zones with populations under 100,000 for the entire period of the sample.[2] Table 3.1 shows the average number of stores of each type across all markets in 1997 and 2007. The data reveals that the number of single-unit stores and small-chain stores decreased significantly. The number of single-unit stores decreases by 29 percent from 2.74 to 1.96. Also, there were fewer small-chain stores between 1997 and 2007, decreasing by 34 percent from 1.04 to 0.69. During the same period, the number of large stores and general merchandisers increased significantly. The number of large stores increased by 12 percent from 2.68 to 2.99. The number of general merchandisers increased by 25 percent from 6.45 to 8.08. Imports also increased significantly. The average market moved from less than 2,000 dollars in clothing imports per worker to more than 7,000 dollars in clothing imports per worker.[3]

There is a significant amount of turnover from year to year. Table 3.2 shows the number of stores of each type that enter and exit in the average year.

On average 0.49 new single-unit stores enter each market each year, but 0.57 exit, which

[2] In particular, the first stage of estimation involves maximization with an expectation-maximization algorithm and 150 parameters in the conditional choice probabilities.

[3] Imports per worker are measured in 2007 U.S. dollars.

Table 3.1: Sample Summary Statistics

	1997	2007
Single-Unit	2.74	1.96
Small Chains	1.04	0.69
Large Firms	2.68	2.99
General Merchandisers	6.45	8.08
Imports	1.07	3.45

Notes: Figures represent the average number of stores of each type of firm and import state for 219 commuting zones for 1997 and 2007.

leads to a decrease in the total number of single-unit stores. Turnover for small chains is lower than for single-unit stores, even after accounting for the smaller number of small-chain stores in each market.

Table 3.2: Number of Stores Entering and Exiting by Type

	Mean	S.D.
Number of new single-units	0.49	0.78
Number of exiting single-units	0.57	0.85
Number of new small-chains	0.09	0.35
Number of exiting small-chains	0.14	0.43
Number of new large stores	0.39	0.72
Number of exiting large stores	0.32	0.64
Number of new general merchandisers	0.89	1.12
Number of exiting general merchandisers	0.72	0.99

Notes: The figures represent the average number of entering and exiting stores of each type per year, across 219 commuting zones and 10 years.

3.3 Model

The previous results establish that increasing competition from direct imports led to the exit of small stores. This section investigates the net effect of direct imports. In particular, it covers how both entry and exit decisions of small retailers depend on competition from direct imports, large retailers, and other small retailers in their markets.

3.3.1 Description of the Model

Overview

The model follows the setup of Arcidiacono, Bayer, Blevins and Ellickson (2016) (ABBE). ABBE use the model to study the response of grocery stores to the entry of Walmart into the grocery business. They model the entry and exit decisions of small stores, large chains, and Walmart. I modify their setup to add an additional store type and direct imports, but only allow each firm to operate one store in a market. This assumption is necessary to preserve tractability and maintain the confidentiality of the firms in the dataset. In the estimation, I will focus on product categories and markets where it is rare for a firm to have multiple stores.

There are four types of stores classified by the size of the firm to which they belong and the industry of the store: single-unit (firms with one store), small chains (firms with 2 to 99 stores), large (firms with more than 100 stores, excluding general merchandisers), and (large) general merchandisers (firms in subsector 452 with more than 100 stores).[4] I will refer to the type of a store by the type of the retailer to which it belongs. The focus is on the entry and exit decisions of single-unit stores and small chains and how they are influenced by competition from large retailers, general merchandisers, and direct imports of competitors. The model takes place in continuous time with players receiving random opportunities to enter or exit.

The key departure from ABBE is to incorporate the degree of direct importing in a market as a state variable. I model the evolution of this state variable in a particular market as depending on the composition of the types of firms in the market. It is important to note that I will be abstracting from national-level importing decisions. Furthermore, for small firms, I will specify an underlying structural model. For large firms, I will estimate reduced form policy functions.

The decisions of small stores depend on the number of stores of each type, the importing decisions of these stores, and market characteristics. I do not formally model the importing decisions of large firms because doing so would make the decisions of large firms interdependent across locations. I assume that small stores track the number of large stores and the average imports of these stores. In the estimation, I allow for these states to depend on each other in a flexible manner.[5]

Timing

Time is continuous and each store receives opportunities to move with rate λ. With each opportunity to move, an incumbent store can elect to exit or do nothing, $j \in \{exit,\ nothing\}$. A potential entrant can elect to enter or do nothing, $j \in \{enter,\ nothing\}$.

There are a number of benefits to setting up the timing in this way instead of having players make simultaneous moves. The primary benefits for this paper are that counterfactuals can be estimated quickly even with a large state space due to the fact that only one agent can move at a time. Thus, from any state only a small number of states are immediately attainable.[6]

[4] This differs from the regression results when general merchandisers were included in both large firms and general merchandisers.

[5] In principal, small stores could also track other moments of the distribution of imports across retailers. The only cost is an increased state space.

[6] ABBE review additional benefits in section 6.4.

Markets

There are many markets that differ in terms of their population (which varies), their average growth rate in population (c), and a permanent unobserved type (z). The growth rate is treated as a permanent observed market type that affects the transition of all states. Thus, agents expect the population to increase at a higher frequency in growing markets.

My treatment of the permanent unobserved type follows Arcidiacono and Miller (2011). The unobserved type is allowed to affect the profitability of entering and remaining in a market differently for stores of every type.[7] It is assumed that each market has a number of potential entrants of each type $\mathcal{E}^h$.[8]

State Space

The state of a market with type (c, z) consists of population level, import level, and number of stores of each type. I define the state x as a vector that contains eight elements. It contains the number of stores by single-unit firms (N^{SU}), small chains (N^C), large firms (N^L), general merchandisers (N^{GM}), the market import penetration level d, and the current population, S:[9]

$$x = (N^{SU}, N^C, N^L, N^{GM}, d, S, c, z).$$

I define the function $l(h, j, x)$ to give the new state conditional on agent h taking action j in state x. For the purposes of this function the agents are $h \in \{SU, C, L, GM, D, S\}$, where D is imports and S is population.

Imports follow a Markov jump process that depends on the state with $F(d'|x)$ the probability that the import state becomes d' at any instant given a current state. Population also follows a Markov jump process and increases with probability $q_u(c)$ and decreases with probability $q_d(c)$.

Flow profits of Single-Unit and Small-Chain Stores

I approximate the flow payoff of a single-unit store, π^{SU}, as a function of market population S, the number of other single-unit stores, N^{SU}, the number of small chain stores, N^C, the number of large stores, N^L, the number of general merchandise stores, N^{GM}, the level of

[7] I assume the market can be in one of five unobserved states $z = \{-1.3998, -0.5319, 0, 0.5319, 1.3998\}$, chosen to be a discretization of a standard normal distribution.

[8] It is assumed that there are three potential single-unit entrants, one potential small chain entrant, three potential large chain entrants, and three potential general merchandiser entrants. These numbers are chosen such that the average entry probability of each type of store is about 10 percent.

[9] d and S are treated as discrete variables.

import penetration, d, and the unobserved state, z. Flow profits for a store of type SU in state x are:

$$\pi^{SU}(x) = \beta_0^{SU} + \beta_{SU}^{SU}(N^{SU}-1) + \beta_C^{SU}N^C + \beta_L^{SU}N^L + \beta_{GM}^{SU}N^{GM} + \beta_d^{SU}d + \beta_S^{SU}S + \beta_T^{SU}\left(N^{SU}\right)^2 + \beta_z^{SU}zN^{SU}$$

I include the square of the number of firms of the same type to allow for economies/diseconomies of scale in the number of single-unit stores in a market.[10] The unobserved state is allowed to affect the degree to which other stores change the profits of small stores.
The flow profits of small chains are defined similarly as:

$$\pi^C(x) = \beta_0^C + \beta_{SU}^C N^{SU} + \beta_C^C(N^C-1) + \beta_L^C N^L + \beta_{GM}^C N^{GM} + \beta_d^C d + \beta_S^C S + \beta_T^C\left(N^C\right)^2 + \beta_z^C zN^C$$

Value Functions of Single-Unit and Small-Chain Stores

For a particular market the value function for store i with firm type $h \in \{SU, C\}$ in state k is given by:

$$\begin{aligned}(\lambda+\rho)V^h(x) = \pi^h &+ \sum_{j\in\{d,u\}} q_j(c)(V^h(l(S,j,x) - V^h(x)) \\ &+ \sum_{d'\in D} F(d'|x)(V^h(l(D,d',x)) - V^h(x)) \\ &+ \sum_{h\in\{S,SC,L,GM\}} \lambda N^h\sigma^h_{exit}(V^h(l(h,exit,x) - V^h(x)) \\ &+ \sum_{h\in\{S,SC,L,GM\}} \lambda\mathcal{E}^h\sigma^h_{enter}(V^h(l(h,enter,x) - V^h(x)) \\ &+ \lambda E\max\{V^h(x) + \varepsilon_{stay}, \varepsilon_{exit}\}\end{aligned} \tag{3.1}$$

The value function depends on the flow profits in that state, the value when the state changes due to population, imports, or a competitor's action, and the value if the agent is allowed to move. At each opportunity to move, stores observe information about the profitability of each choice in terms of an instantaneous payoff ε_j that is unobserved to the econometrician. I assume that these payoffs are distributed independently over time according to a type 1 extreme value distribution.[11]

[10] For example, initially single-unit stores could share distribution networks, lowering their costs.
[11] If a store closes, it is assumed that it cannot reopen, which implies the value of exit is zero.

Choice Probabilities

A potential entrant that receives an opportunity to move must pay a fixed cost to build a store that depends on the type of firm, f^h, and the unobserved state. Thus, given some current state x a firm will enter if

$$V^h(l(h, enter, x)) + f^h + \gamma^h z + \varepsilon_{enter} \geq \varepsilon_{stay},$$

The probability that an entrant of type h enters is

$$\sigma^h_{enter}(x) = \frac{\exp(V^h(l(h, enter, x)) + f^h + \gamma^h z)}{\exp(V^h(l(h, enter, x)) + f^h + \gamma^h z) + 1}.$$

Similarly, an incumbent will exit if

$$V^h(x) + \varepsilon_{stay} \leq \varepsilon_{exit}.$$

Then the probability that an incumbent of type h exits is

$$\sigma^h_{exit}(x) = \frac{1}{\exp(V^h(x)) + 1}.$$

Equilibrium

I focus on Markov perfect equilibria in pure strategies as characterized by Aguirregabiria and Mira (2007). Thus, a Markov perfect equilibrium is a collection of policy functions $\sigma = \{\sigma^{SU}_{enter}, \sigma^{C}_{enter}, \sigma^{L}_{enter}, \sigma^{GM}_{enter}, \sigma^{SU}_{exit}, \sigma^{C}_{exit}, \sigma^{L}_{exit}, \sigma^{GM}_{exit}\}$ such that σ^h_{enter} and σ^h_{exit} solve the maximization problems of each type h.

3.3.2 Estimation

I estimate the model for stores that are in the clothing subsector (448). Clothing is a department characterized by many large firms, high direct imports, and significant competition from general merchandisers. General merchandisers account for about 40 percent of clothing sales in 2007 (CRT, 2007). Clothing is the second largest department, behind groceries, making it the largest department with significant exposure to trade.[12]

[12]Estimation is in progress with electronics and appliances as well.

Estimation Algorithm

Estimation takes place in two steps. In the first step, I estimate conditional choice probabilities and the probability that each market is in unobserved state z_k. In the second step, I estimate the flow payoff and fixed costs of entry parameters of single-unit and small-chain stores taking the conditional choice probabilities of the other stores as given.

Conditional Choice Probabilities

In the first step, I estimate the probability that player i makes choice j in state k with unobserved state z. I assume

$$\sigma_{ij}(k, z, \alpha) = \frac{\phi_j(k, z, \alpha)}{\sum_{j' \in A_{ik}} \phi_{j'}(k, z, \alpha)},$$

where ϕ is a function of: a constant; the number of single-unit establishments and their square; the number of small chain establishments and their square; the number of large establishments and their square; the number of general merchandise establishments and their square; the square of the number of total establishments; the level of import competition; indicators of the market type; the unobserved state; and the interaction of population with the number of stores of each type. Additionally, I allow the import penetration transition probabilities to depend on the share of large stores and general merchandise stores and the fixed cost of building a store to depend on the unobserved state. For computational simplicity, the transition probability of the population is estimated using the frequency of population transitions between 1997 and 2007 in the markets used in the sample.

Objective Function

In many markets, there are multiple openings and closings within a year. In the data there is only one observation per year, but the model takes place in continuous time so I simulate R paths for each observation, which consist of sequences and timings for each of the M moves that took place in the market during the year.[13]

I define the likelihood of a single observation n in market m, where the starting and ending states are $\underline{k}$ and $\bar{k}$. Let W be the number of events that occurred during the year. Let $k_w^{(r)}$ denote the state immediately preceding event w in simulation r, with $w = 1, \ldots, W+1$. I simulate paths $r = 1, \ldots, R$ such that $k_1^{(r)} = \underline{k}$ and $k_{W+1}^{(r)} = \bar{k}$. Let $I_w^{(r)}(i, j)$ be the indicator

[13] I do not observe the specific time during the year during which each establishment entered or exited. ABBE find similar estimates using exact Walmart entry dates as when they use only yearly information.

for whether event w of the r-th simulation was action j taken by firm i and let $t_w^{(r)}$ and $\tau_w^{(r)}$ be the absolute time and holding time of simulated event w.

The likelihood for observation n in market m is

$$\begin{aligned}
\tilde{L}_{mn}(h(\alpha); z) = \frac{1}{R} \sum_{r=1}^{R} \prod_{w=1}^{W} & \left(\sum_{j \in \{-1,1\}} I_w^{(r)}(0,j) q_j + \sum_i \lambda \sum_{j \neq 0} I_w^{(r)}(i,j) \tilde{\sigma}_{ij} \left(k_w^{(r)}, z, \alpha \right) \right) \\
& \times \exp \left[- \left(\sum_{j \in \{-1,1\}} q_j + \sum_i \lambda \sum_{j \neq 0} \tilde{\sigma}(k_w^{(r)}, z, \alpha) \right) \tau_w^{(r)} \right] \\
& \times \exp \left[- \left(\sum_{j \in \{-1,1\}} q_j + \sum_i \lambda \sum_{j \neq 0} \tilde{\sigma}_{ij}(k_{W+1}^{(r)}, z, \alpha) \right) \left(1 - t_W^{(r)} \right) \right].
\end{aligned} \tag{3.2}$$

The first line of equation (3.2) is the probability that event w occurred, the second is the probability that no other event occurred during time period $\tau_w^{(r)}$, the final line is the probability that no event occurred between the last simulated event and the end of the period.

Unobserved Heterogeneity:

Since z is unobserved, I estimate the probability each market has type z_k as a function of initial conditions.[14] I allow z_k to take five values which are chosen to approximate a standard normal distribution.

Let $P(z, k_1)$ be the probability of the unobserved state being z, given that the observed state was k_1 for the first observation. With M markets and T periods in each, summing with respect to the distribution of the unobserved state yields

$$(\tilde{\alpha}, \tilde{P}) = \arg \max_{(\alpha, P)} \sum_{m=1}^{M} \ln \left(\sum_z P(z, k_{m1}) \prod_{n=1}^{T} \tilde{L}_{mn}(h(\alpha); z) \right). \tag{3.3}$$

This is estimated using the Expectation Maximization (EM) algorithm following Arcidiacono and Miller (2011) to get both reduced form hazards and the probability each market is in an unobserved state.

[14]Specifically, the probability that a market has a particular value of the unobserved state is modeled as an ordered logit that depends on the number of stores of each type, the interaction of these counts with population, the import state, and the city growth time in the initial period.

Estimation of Structural Parameters

Given the estimates of the reduced form hazard function, I can turn to estimating the structural parameters of each firm.

From proposition 4 in ABBE, I can rewrite the value function in terms of the estimated choice probabilities and structural parameters. Then, estimating the structural parameters consists of maximum likelihood estimation as a function of the structural parameters.

3.3.3 Effects of Trade in the Model

I use my model to conduct a counterfactual to study the effects of trade in the retail sector. The exercise seeks to compare the economy exposed to trade with China (the U.S. economy after 1997) with a counterfactual economy that was not exposed to the 1997 China shock. The difficulty in this exercise is estimating how the large stores and general merchandisers would respond to this alternative policy change.

I introduce an additional parameter, τ, which specifies the trade regime. In particular, τ^0 is the trade regime prior to 1997 when very few imports came from China and τ' is the trade regime after the 1997 China shock. The policy functions estimated using data between 1997 and 2007 were estimated under the trade regime τ'. So the estimated policy functions are $\sigma^h_{enter}(x, \tau'), \sigma^h_{exit}(x, \tau')$ for $h \in \{SU, C, L, GM\}$. I assume that τ affects the transition probability on the import state such that a higher τ implies the import state is more likely to be high. Formally, let $\tau' > \tau$, then $F(d|x, \tau') < F(d|x, \tau)$.

I place two assumptions on the policy functions of large stores and general merchandisers which allow me to calculate a lower bound on the effect of trade:

1. Entry of large stores and general merchandisers is increasing in the trade regime

$$\sigma^h_{enter}(x, \tau^0) \leq \sigma^h_{enter}(x, \tau') \text{ for } h \in \{L, GM\}.$$

2. Entry of large stores and general merchandisers is increasing in the import state

$$\frac{\partial \sigma^h_{enter}(x, \tau')}{\partial d} > 0.$$

The first assumption is consistent with estimates of Meinen and Raff (2018) and the second assumption is consistent with the estimated policy functions.

These assumptions imply

$$\tilde{\sigma}^h_{enter}(x, \tau^0) = \sigma^h_{enter}(x, \tau')$$

for $h \in \{L, GM\}$ is a lower bound on the effect of moving from τ' to τ^0. In this scenario, the behavior conditional on x does not change, but under τ^0 it is assumed that $d = 0$. Given these policy functions for large stores and general merchandisers, I calculate the best responses for single-unit and small-chain stores which I refer to as $\tilde{\sigma}^h_{enter}$ and $\tilde{\sigma}^h_{exit}$. Under assumptions 1 and 2, the share of small stores under τ^0 is less than the share calculated using the policy functions $\tilde{\sigma}^h_{enter}(x, \tau^0)$. Thus, this counterfactual provides a lower bound on the effect of direct imports.

Let μ be the stationary distribution of x given some trade regime τ. The counterfactual involves comparing μ' to μ^0 to calculate the stationary share of single-unit and small-chain stores.

3.4 Results

Table 3.3 reports profit coefficients for single-unit clothing stores and small-chain clothing stores both with and without unobserved heterogeneity. Imports and competition with general merchandisers negatively affect the profits of both types of small stores. Surprisingly, single-unit stores are more affected by competition with both imports and general merchandisers. Consider a single-unit store that has an entry probability of 50 percent. The coefficient on imports of -0.57 implies that an increase in the import state lowers the entry probability to 36 percent. The coefficient of -0.27 on general merchandisers suggests the entry probability would decrease to 43 percent. Both of those are significant effects given that the level of imports increased by 1.38 on average and the number of general merchandisers increased by 1.63.

The effect of large stores on profits is more mixed. Without unobserved heterogeneity I find both single-unit stores and small chains are both negatively affected by competition with large firms, but when unobserved heterogeneity is included, no effect of large stores on single units is found. I find evidence of economies of scale with small stores at first, but these fall off as the number of stores increases. The unobserved heterogeneity results suggest positive returns to scale until 7 single-unit stores and 6 small chains are present when the unobserved state is zero. In all cases, I find fixed costs are larger for small chains than single units, but the unserved state decreases the fixed cost of small chains significantly. On the other hand, higher unobserved states decreases the profits of single-unit stores.

Table 3.3: Structural Parameter Estimates

	No Unobserved Heterogeneity		Unobserved Heterogeneity	
	SU	C	SU	C
Constant (β_0)	−20.370	−22.370	−16.870	−13.440
Number of Single-Unit Stores (β_{SU})	1.501	−0.095	1.008	−0.186
Number of Small-Chain Stores (β_C)	−0.159	2.482	−0.254	1.967
Number of Large Stores (β_L)	−0.318	−0.192	0.022	−0.185
Number of GM Stores (β_{GM})	−0.582	−0.230	−0.276	−0.064
Import Penetration (β_d)	−0.625	−0.401	−0.570	−0.080
Population (β_S)	0.661	1.156	0.528	0.079
Number of own type squared (β_T)	−0.174	−0.305	−0.070	−0.143
Unobserved state×number of own type (β_z)			−0.119	−0.231
Entry cost (f)	−1.780	−3.639	−2.010	−6.434
Entry cost × unobserved state			−0.063	2.964

Notes: Estimates of structural parameters for stores in the clothing industry. Coefficients represent the effect of each variable on flow profits by type of store. SU - Single-unit, C-Small Chain. Coefficients in value function units. Unobserved heterogeneity specifications include 5 unobserved states.

3.5 Counterfactual: Retail Without Direct Imports

In this section, I run a series of counterfactuals where imports are removed to estimate their effect on the clothing industry. Removing imports has two effects. First, it increases flow profits of small stores making them more likely to import and less likely to exit. Second, it changes the entry and exit decisions of large stores both through a direct effect on their profits and through their response to the different policy functions of the other firms.

I use my estimates of structural parameters for the small stores to remove direct imports from the flow profits of the small stores. Then, I simulate the model with different assumptions regarding how the policy functions of large stores and general merchandisers would change if imports were removed. This simulation provides estimates of how removing imports would affect the structure of retail. The estimates depend on the direct effect of eliminating imports and different assumptions on the indirect effect.

In each simulation, I compare the number of stores of each type in 2007 to the number in the baseline case. I also compute the average of the local HHI. It is assumed that all stores of each type have sales equal to the average store of that type in the sample and that each firm operates only one shop.[15] All values are taken from the Census of Retail Trade in 2007. In all scenarios, I modify the relevant parameters of the conditional choice probabilities and structural parameters. Then, I use value function iteration to solve for new best responses of both types of small stores.[16]

No Imports in Flow Profits

The first counterfactual considered is one where direct imports do not affect the flow profits of small stores, but the policy functions of large stores and general merchandisers remain unchanged. I simulate each market's state in 2007 using 1997 as the starting point with $\beta_I^s = \beta_I^c = 0$. I compare these results to the baseline in table 3.4. I find that removing imports from the flow profits increases the number of both single-unit stores and small chains, but the effects are relatively small. The number of single-unit stores increased by eight percent and the number of small chains increased by almost three percent. Together these imply the total number of small stores increased by six percent. The number of small clothing stores decreased by 30 percent during this period, which suggests direct imports account for at least 14 percent of the decrease. I calculate the average HHI across all markets using the average value of clothing sales for stores of each type. Although the average number of small stores increased, I find no change in the average HHI across markets.

Entry of Large Firms

In the second counterfactual, I set $d = 0$ and solve for new best responses of small stores. This counterfactual provides a tighter lower bound on the effect of importing under the assumptions in section 3.3.3.[17]

---

[15] The markets used are small enough it is rare for a firm to have multiple stores in a market.

[16] The resulting policy functions should not be thought of as equilibrium policy functions because the modified conditional choice probabilities I use for large and general merchandise stores may not be equilibrium strategies.

[17] Estimation of this counterfactual is in process.

Table 3.4: Counterfactual Results

	No Unobserved Heterogeneity				
	Single Unit	Small Chain	Large	GM	Average HHI
Baseline	3.203	2.612	9.528	6.094	0.08
$\beta_I^s = \beta_I^c = 0$	3.469	2.684	8.934	5.813	0.08

Notes: Estimates of counterfactual number of stores of each type and local HHI. Rows are for different assumptions on how the policy functions of large firms change when direct imports are removed.

3.6 Conclusion

I develop a model of retailer competition with direct imports to test how competition with direct imports affects the entry and exit decisions of small stores. I find that direct imports account for at least 14 percent of the decrease in the number of small stores. The results of this research indicate that direct imports have played an important role in increasing national concentration in the retail sector, but have little effect on local concentration.

Bibliography

Aguirregabiria, Victor and Pedro Mira (2007) "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, Vol. 75, pp. 1–53.

Aguirregabiria, Victor and Aviv Nevo (2011) "Recent developments in empirical IO: Dynamic demand and dynamic games," *Advances in Economics and Econometrics: Tenth World Congress Volume 3, Econometrics*, pp. 53–122.

Akcigit, Ufuk and Sina T Ates (2019) "Ten Facts on Declining Business Dynamism and Lesssons from Endogeneous Growth Theory," `http://www.nber.org/papers/w25755`.

Amiti, Mary, Mi Dai, Robert C Feenstra, and John Romalis (2017) "How Did China's WTO Entry Affect U.S. Prices?," `https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr817.pdf?la=en`.

Antras, Pol, Teresa C Fort, and Felix Tintelnot (2017) "The Margins of Global Sourcing: Theory and Evidence from U.S. Firms," *American Economic Review*, Vol. 107, pp. 2514–2564.

Arcidiacono, Peter, Patrick Bayer, Jason R Blevins, and Paul B Ellickson (2016) "Estimation of Dynamic Discrete Choice Models in Continuous Time with an Application to Retail Competition," *Review of Economic Studies*, Vol. 83, pp. 889–931.

Arcidiacono, Peter and Robert A Miller (2011) "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, Vol. 79, pp. 1823–1867.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen (2017) "The Fall of the Labor Share and the Rise of Superstar Firms."

Basker, Emek (2005) "Job Creation or Destruction? Labor Market Effects of Wal-Mart Expansion," *The Review of Economics and Statistics*, Vol. 87, pp. 174–183.

Basker, Emek, Shawn Klimek, and Pham Hoang Van (2012) “Supersize It: The Growth of Retail Chains and the Rise of the ”Big-Box” Store,” *Journal of Economics and Management Strategy*, Vol. 21, pp. 541–582.

Basker, Emek and Michael Noel (2009) “The Evolving Food Chain : Competitive Effects of Wal-Mart’s Entry Into,” *Journal of Economics and Management Strategy*, Vol. 18, pp. 977–1009.

Basker, Emek and Pham Hoang Van (2010) “Imports “R” Us : Retail Chains as Platforms for Developing-Country Imports,” *American Economic Review: Papers & Proceedings*, Vol. 100, pp. 414–418.

Bernard, Andrew B, Bradford J Jensen, Stephen J Redding, and Peter K Schott (2018) “Global Firms,” *Journal of Economic Literature*, Vol. 56, pp. 565–619.

Bernard, Andrew B, J Bradford Jensen, Stephen J Redding, and Peter K Schott (2010) “Wholesalers and Retailers in US Trade,” *American Economic Review: Papers & Proceedings*, Vol. 100, pp. 408–413.

Davis, Steven J., John Haltiwanger, and Scott Schuh (1996) “Small Business and Job Creation: Dissecting the Myth and Reassessing the Facts,” *Small Business Economics*, Vol. 8, pp. 297–315.

De Loecker, Jan and Jan Eeckhout (2017) “The Rise of Market Power and the Macroeconomic Implications,” `http://www.nber.org/papers/w23687`.

Decker, Ryan A, John Haltiwanger, Ron S Jarmin, and Javier Miranda (2017) “Declining Dynamism, Allocative Efficiency, and the Productivity Slowdown,” *American Economic Review: Papers & Proceedings*, Vol. 107, pp. 322–326.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda (2014) “The Role of Entrepreneurship in US Job Creation and Economic Dynamism,” *Journal of Economic Perspectives*, Vol. 28, pp. 3–24.

Fort, Teresa C and Shawn D Klimek (2016) “The Effects of Industry Classification Changes on US Employment Composition,” `http://faculty.tuck.dartmouth.edu/images/uploads/faculty/teresa-fort/fort_klimek_naics.pdf`.

Foster, Lucia, John Haltiwanger, Shawn Klimek, C.J. Krizan, and Scott Ohlmacher (2015) “The Evolution of National Retail Chains: How We Got Here,” `http://ssrn.com/abstract=2589175`.

Ganapati, Sharat (2018) "The Modern Wholesaler: Global Sourcing, Domestic Distribution, and Scale Economies."

Hall, Robert E (2018) "New Evidence of the Markup of Prices Over Marginal Costs and the Role of Mega-Firms in the US Economy," `http://www.nber.org/papers/w24574`.

Haltiwanger, John C, Ron S Jarmin, and C. J. Krizan (2010) "Mom-and-Pop Meet Big-Box: Complements or Substitutes?," *Journal of Urban Economics*, Vol. 67, pp. 116–134.

Handley, Kyle and Nuno Limão (2017) "Policy Uncertainty, Trade, and Welfare: Theory and Evidence for China and the United States," *American Economic Review*, Vol. 107, pp. 2731–2783.

Holmes, Thomas J (2011) "The Diffusion of Wal-Mart and Economies of Density," *Econometrica*, Vol. 79, pp. 253–302.

Holmes, Thomas J and Ethan Singer (2018) "Indivisibilities in Distribution," `http://www.nber.org/papers/w24525`.

Hortaçsu, Ali and Chad Syverson (2015) "The Ongoing Evolution of US Retail: A Format Tug-of-War," *Journal of Economic Perspectives*, Vol. 29, pp. 89–112.

Hummels, David, Rasmus Jørgensen, Jakob Munch, and Chong Xiang (2014) "The Wage Effects of Offshoring: Evidence from Danish Matched Worker-Firm Data," *American Economic Review*, Vol. 104, pp. 1597–1629.

Jaravel, Xavier and Erick Sager (2018) "What are the Price Effects of Trade? Evidence from the U.S. and Implications for Quantitative Trade Models," `https://docs.wixstatic.com/ugd/bacd2d_8cad0fefc5ae4a1886ee58075019c431.pdf`.

Jia, Panle (2008) "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry," *Econometrica*, Vol. 76, pp. 1263–1316.

Justice, Department of and Federal Trade Commision (2010) "Horizontal Merger Guidelines."

Lipsius, Ben (2018) "Labor Market Concentration Does Not Explain the Falling Labor Share," `https://www.dropbox.com/s/edlepa1we8vp6ye/lipsius_JMP.pdf?dl=0`.

Meinen, Philipp and Horst Raff (2018) "International Trade and Retail Market Performance and Structure: Theory and Empirical Evidence," *Journal of International Economics*, Vol. 115, pp. 99–114.

Rinz, Kevin (2018) "Labor Market Concentration, Earnings Inequality, and Earnings Mobility," `https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/carra-wp-2018-10.pdf`.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Nicholas Trachter (2019) "Diverging Trends in National and Local Concentration."

Smith, Dominic (2019) "Foreign Sourcing in the U.S. Retail Sector."

Smith, Dominic and Sergio Ocampo (2019) "The Evolution of U.S. Retail Concentration."

Appendix A

Appendix to Chapter 1

A.1 Comparison to Rossi-Hansberg, Sarte and Trachter (2019)

Prior papers studying competition in the retail sectors have defined each store's competitors using industry codes. In particular, a recent paper by Rossi-Hansberg, Sarte and Trachter (2019) find that despite increases in aggregate concentration at the SIC 8-digit, level local concentration has fallen across many geographies for many sectors. In this section, we compare my results to their results on the retail sector and speculate on the reason for differences.

As mentioned in Section 1.4, there are three main differences between our work and RST's. First, we use different data sources. While they use the U.S. NETS dataset we use confidential Census microdata. Second, we use NAICS codes to define industries in the retail sector, while they use SIC codes. Third, RST report the average change in the local HHI, weighting by the end-of-period sales/employment of each market, while we report the change in the average local HHI, weighting markets in each year according to that year's sales. To better understand the differences between our work and RST's, we present industry- and product-based measures of concentration varying the weighting scheme used in aggregation. Table A.1 has the results of the comparison. In the first section, national concentration, we compare the numbers in RST (Figure 1b) to numbers calculated for three different samples. NAICS-based measures calculate concentration separately for all 6-digit industries in NAICS. Select NAICS calculates concentration for all 6-digit NAICS excluding auto dealers and auto-parts stores (441), gasoline stations (447), and non-store retailers (454). Finally, product-based measures calculate concentration for the eight major departments defined in Section 1.5. In all four cases, national concentration is increasing significantly. Despite differences in the initial levels of concentration (column 1) the national HHI increases by two to three times in all cases.[1]

Table A.1: Comparison of Concentration to RST

National Concentration				
	Level	Change from 1992		
	1992	1997	2002	2007
RST	N/A	0.020	0.030	0.050
NAICS-based	0.029	0.017	0.056	0.076
Select NAICS	0.046	0.034	0.097	0.136
Product-based	0.015	0.006	0.027	0.040
Zip Code Concentration - End-of-Period Weights				
	Level	Change from 1992		
RST	N/A	-0.070	-0.100	-0.140
NAICS-based	0.507	0.024	-0.018	-0.019
Select NAICS	0.552	-0.021	-0.018	-0.015
Product-based	N/A	N/A	N/A	N/A
Zip Code Concentration - Current Period Weights				
	Level	Change from 1992		
RST	N/A	N/A	N/A	N/A
NAICS-based	0.507	0.022	0.057	0.072
Select NAICS	0.552	0.026	0.067	0.083
Product-based	0.321	-0.015	0.020	0.033

Notes: Comparison of concentration numbers calculated using the Census of Retail Trade to Rossi-Hansberg et al. (2019). Numbers from RST taken from retail series in Figure 1b. 1992 column contains the level of concentration which is not available in RST. NAICS-based measures concentration calculated including all NAICS industries. Select NAICS drops subsectors 441, 447, and 454. Product-based measures calculate concentration for the eight major product categories. Retail in RST is defined using SIC codes which includes restaurants. Product-based measures with RST's methodology have not been disclosed.

The second portion of table A.1 compares concentration measured at the zip code level using the weighting methodology described in RST. In particular, RST calculate the change in concentration for each market-industry pair between 1992 and some year t. Those changes are aggregated to an index of local concentration using the share of employment in each industry-market pair *in year* t. Using the methodology in RST we find evidence for slight

[1]The level of concentration is not provided in RST.

decreases in local concentration. RST find local concentration falls by 14 percentage points, but we find it falls by less than two percentage points. These results partially reconcile our findings with RST's, with differences remaining in the type of industry classification and data source.

The final part of table A.1 compares concentration measured at the zip code level by calculating concentration for each market-industry pair within a year then aggregating to an index of local concentration for each year. These aggregate local indexes are then differenced to calculate the change in concentration. We find evidence for increasing concentration. Importantly, this happens at both the NAICS and department level, suggesting that the difference in the direction of the trends of local concentration is due to the weighting methodology and not the use of of industry- or product-based measures.

The results under the two weighting schemes (panels 2 and 3) differ for a simple reason, when markets grow they tend to become less concentrated. Thus, weighting changes in concentration using period t shares weights markets with decreases in concentration more than if one uses current period shares. However, weighting using current period shares is not without problems. In particular, local concentration can change only because of changes in the weights on each market. We use current period shares when calculating concentration because those are the weights for which my decomposition holds.

A.2 Concentration Decomposition

This section provides additional details on the concentration decomposition.

The decomposition starts from the probability that two dollars are spent at the same firm which can be divided into three components.

$$P(i_x = i_y) = P(i_x = i_y | m_x = m_y) P(m_x = m_y) + P(i_x = i_y | m_x \neq m_y) P(m_x \neq m_y)$$

The local concentration term can be decomposed into a component due to the average number of firms in each market and a term due to the inequality of shares across firms within a market.

$$
\begin{aligned}
P\left(i_x = i_y | m_x = m_y\right) &= \sum_{m=1}^{M} \underbrace{P\left(m_x = m | m_x = m_y\right)} \overbrace{P\left(i_x = i_y | m_x = m, m_x = m_y\right)}^{\text{Local HHI}} \\
&= \sum_{m=1}^{M} \frac{s_m^2}{\sum_n s_n^2} \sum_{k=1}^{K} \left(s_k^m\right)^2 \\
&= \sum_{m=1}^{M} s_m \left(\frac{1}{N_m} + \sum_{k \in K_m} \left(s_k^m - \frac{1}{N_m}\right)^2\right) \\
&= \underbrace{\sum_{m=1}^{M} s_m \frac{1}{N_m}}_{\text{Average Number of Firms}} + \sum_{m=1}^{M} s_m \sum_{i \in K_m}^{K} \underbrace{\left(s_k^m - \frac{1}{N_m}\right)^2}_{\text{Inequality of shares}}
\end{aligned}
$$

The cross market term is defined as

$$
P\left(i_x = i_y | m_x \neq m_y\right) = \sum_k \sum_{j \neq k} P\left(m_x = k, m_y = j | m_x \neq m_y\right) P\left(i_x = i_y | m_x = k, m_y = j\right)
$$

$$
P\left(m_x = k, m_y = j | m_x \neq m_y\right) = \frac{s_k s_j}{1 - \sum_m s_m^2}
$$

$$
P\left(i_x = i_y | m_x = k \wedge m_y = j\right) = \sum_{i=1}^{I} s_{ik} s_{ij}
$$

$$
\begin{aligned}
\left(1 - P\left(m_x = m_y\right)\right) P\left(i_x = i_y | m_x \neq m_y\right) &= \left(1 - \sum_m s_m^2\right) \sum_k \sum_{j \neq k} \frac{s_k s_j}{1 - \sum_m \left(s_m^N\right)^2} \sum_{i=1}^{I} s_{ik} s_{ij} \\
&= \sum_k \sum_{j \neq k} s_k s_j \sum_{i=1}^{I} s_{ik} s_{ij}.
\end{aligned}
$$

Appendix B

Appendix to Chapter 2

B.1 Samples and Data Methods

B.1.1 Empirical Sample Selection Criteria

Selection into the empirical sample is based on both CRT and LBD activity. I calculate the percentage of each firm's employment that is in the retail sector using the LBD. I define a firm to be a retail firm if at least half of its employees are in retail establishments.[1] I require a retail firm to have at least one establishment included in official tabulations.[2] If a firm has no employment in the LBD, but has sales in the Census of Retail Trade it is included in the sample. I define the size of a firm based on the number of establishments it has in the retail sector.

I assign each firm a subsector (3-digit NAICS) based on the plurality of their employment in retail. Retail firms typically have establishments in only one subsector. The average firm has 99 percent of its employment in the subsector to which it is assigned, and over 95 percent of retail employment is in the top subsector of the firm that owns the establishment. I assign establishments to commuting zones using the concordance from David Dorn.

B.1.2 Cleaning and Aggregating Product Lines Data

The Census collects data on establishment-level sales in a number of categories of goods. An example form is provided in figure B.1. Many establishments have missing product line sales either due to them not responding to questions or because they do not receive a form.[3]

[1] This cutoff does not matter much. Firms typically have either more than 90 percent of their employment in retail or almost none of their employment in retail.

[2] This means tabbed is yes and the new NAICS code assigned to it is retail for 1992 and later. I also drop establishments in Hawaii and Alaska as well as those that are missing geographic information.

[3] Establishments of large firms are always mailed a form, but small firms are sampled.

In total reported product lines, data account for about 80 percent of sales. I develop an algorithm to impute data for missing establishments. This involves aggregating product line codes to the point that the industry of an establishment and the establishment's answer to the kind of business question is highly informative about their sales. For example, I aggregate lines for Women's clothes, Men's clothes, children's clothes, and footwear into a department called clothing. I establish 20 departments detailed in table B.1.2. Of these 20 departments, 8 of them account for the vast majority of sales of stores in my sample. The other 12 departments are specialty categories that account for a small fraction of aggregate sales and are sold primarily by establishments in one specific industry. For example, glasses are sold almost exclusively by establishments in 446130 (optical goods stores). I create these categories so that establishments that sell these products are not included in concentration measures for the 8 main departments.

Aggregating Product Lines

The first step of cleaning the data is to aggregate reported broad and detailed product line codes into departments. Some codes reported by retailers do not correspond to valid product line codes. I allocate those sales to a miscellaneous department. The Census analyzes reported product line codes to check for issues and flags observations as usable if they pass this check. I include only observations that are usable. I then map these codes to departments. I use the reported percentage of total sales accounted for by each product line instead of the dollar value because the dollar value is often missing. Typically an establishment either reports product line data for 100 percent of its sales or does not report any data. For the small number of establishments that report product lines data summing to a number other than 100 percent I rescale the percentages so that they sum to one.[4] After this procedure, I have sales by department for all establishments that reported lines data.

Imputing Missing Data

For the remaining establishments I impute data using the NAICS 8 industry of the establishment (where available), reported sales of other establishments of the same firm in the same industry, and reported activity of the same establishment in other census years.[5] Most establishments are assigned a distribution of sales across departments that matches

[4] This procedure has a minimal effect on aggregate retail sales in each department.

[5] Reported product line sales are very similar across establishments of the same firm and the same establishment over time.

the mean of establishments in the same industry that report sales.

I find that this procedure predicts sales accurately for most establishments, but a small number of stores in each industry report selling very different things than all other stores in that industry. In these cases, the prediction misses by a lot.

Using this aggregation method, establishments overwhelmingly only have significant sales in two departments which increases confidence in the imputation. Additionally, I have compared the aggregate sales in my data to the Consumer Expenditure Survey which is an independent program, and they are in line with the numbers from that source.[6]

Where relevant all sales are deflated using consumer price indexes. I use the food deflator for Groceries, Clothing and Apparel deflator for Clothing and the deflator for all goods excluding food and fuel for all other categories.

Table B.1: List of Departments

Department	Main	Corresponding Industry	Example Firm
Automotive Goods	N	441	Ford Dealer
Clothing	Y	448	Old Navy
Electronics and Appliances	Y	443	Best Buy
Furniture	Y	442	Ikea
Services	N	N/A	
Other Retail Goods	N	N/A	
Groceries	Y	445	Trader Joe's
Health Products	Y	446	CVS
Fuel	N	447	Shell Gasoline
Sporting Goods	Y	451	Dick's Sporting Goods
Toys	Y	451	Toys "R" Us
Home & Garden	Y	444	Home Depot
Paper Products	N	453210	
Jewelry	N	423940	Jared
Luggage	N	448320	Samsonite
Optical Goods	N	446130	Lenscrafters
Non-retail Goods	N	N/A	
Books	N	451211	Borders

Notes: Author created list of departments. Main indicates that a department is included in concentration calculations. Firm names for illustrative purposes only and do not imply that firm is in the analytical sample.

[6] Retail sales include some sales to companies so it is expected that retail sales in a department to exceed consumer spending on that department.

B.1.3 Mapping HS Codes to Product Lines

The starting point from my mapping of HS codes to product lines is the concordance from Basker and Van (2010) which maps HS codes to broad product lines. I then aggregate broad product lines to departments using the same concordance as in the product lines data. I then update the concordance using the LFTTD and CRT. Specifically, I create a list of HS codes that are imported by retail firms, but not assigned to a product line by Basker and Van (2010). This approach means that almost all retailer imports are defined to be consumer goods imports, but can overstate total imports of consumer goods.[7] I then assign these codes to departments by investigating the codes that have the most imports by retailers manually and then assigning the rest using the industry of the firms that do the importing.

Using this procedure ensures retailer imports are almost all mapped to departments, but it can cause me to overstate total imports of consumer goods to the extent there are HS codes that are imported by retailers for sale to consumers, but are also imported by other firms where the final customer is not an individual, but a business. Ideally, I would identify all imports that end up in consumer's hands without any transformation and then see what fraction are imported directly by retailers, but this is not possible. Instead, I use a procedure that gives me a lower bound on the fraction of consumer goods imported directly by retailers.

I deflate imports using the BLS import deflators for the 20 HS sections.

B.1.4 Model Sample Selection Criteria

Estimation of the model requires only information on the existence of establishments, their industry, firm type, and competition from imports. Thus, I use only the LBD as that lets me observe establishments on a yearly basis. Establishments can change their firm type and industry over time so I use the mode of their size category and industry between 1991 and 2008. Most establishments do not change type or industry during this period.

I calculate the number of general merchandise establishments and firms as the number of stores in 452 excluding NAICS 452990 (Other General Merchandise Stores), because those stores do not report selling a significant amount of clothing.

I calculate commuting zone population using data downloaded from `http://data.nber.org/data/census-intercensal-county-population.html`. The data are county level population estimates by year. I aggregate to commuting zones using the concordance from

[7]For example a retailer may import a desk to sell to consumers, but other companies import desks to furnish offices.

David Dorn. I take the log of population and create 7 categories of population such that each population level represents an increase of 20 percent from the previous level.
Commuting zone population growth rate is divided into three categories and is permanent. The category is fast if the annual growth rate is over 2 percent, moderate if it is between 1 and 2 percent, and low if it is less than 1 percent.
I cannot measure exposure to imports as in the regressions because I only observe sales every 5 years. Instead I measure firm level imports by department and assign them to individual establishments using that establishment's share of firm employment. I sum across all imports in a department assigned to a given commuting zone and divide those imports by total employment in the industry most closely corresponding to the imports.

B.1.5 Additional Regression Results

I present results with an additional control for whether the store already imported at least one percent of its sales directly in 2002. Table B.2 shows that single-unit stores that import are actually much more likely to exit. Stores that imported in 2002 are 5.9 percentage points more likely to exit. However, for both types of small stores importers that survive grow more than the surviving importers. These results suggest that whether a small store imports may be a signal that the small store sells foreign made products instead of being a niche retailer. Thus, the small store may face more competition on the fraction of its sales that are not imported directly.

Figure B.1: Sample Product Lines Form

Item 10. MERCHANDISE LINES

Report sales for each merchandise line sold by this establishment, either as a dollar figure or as a whole percent of total sales. *(See HOW TO REPORT DOLLAR FIGURES on page 1 and HOW TO REPORT PERCENTS below)*

HOW TO REPORT PERCENTS	If figure is **38.76%** of total sales:	Mil.	Thou.	Dol.	Percent
	• ***Report whole percents*** →				**39**
	Not acceptable →				38.76

Merchandise lines	Census use	ESTIMATES are acceptable. Report dollars OR percents. Mil.	Thou.	Dol.	Percent
	230	231			232
1. Women's, juniors', and misses' wear (Report girls' and infants' and toddlers' wear on line 3 and footwear on line 4)	**0220**				
2. Men's wear (Report boys' wear on line 3 and footwear on line 4)	**0200**				
3. Children's wear (Include boys' (sizes 2 to 7 and 8 to 20), girls' (sizes 4 to 6x and 7 to 14), and infants' and toddlers' clothing and accessories. Report footwear on line 4.)	**0240**				
4. Footwear (include accessories)	**0260**				

FORM RT-5302

Table B.2: Importing and Small Store Outcomes

	Single-Unit		Small Chain	
Dependent variable is:	Exit	Pct Change	Exit	Pct Change
$\Delta X_{im}^{2002-2007}$	0.755**	1.569*	1.701**	-2.636
	(0.319)	(0.853)	(0.809)	(1.989)
X_{im}^{2002}	0.456**	1.166***	1.189**	0.694
	(0.228)	(0.452)	(0.462)	(1.418)
pct_{im}^{L}	0.125***	-0.183**	0.104***	-0.185***
	(0.012)	(0.023)	(0.027)	(0.047)
pct_{im}^{GM}	-0.104***	-0.123***	0.160***	0.015
	(0.018)	(0.029)	(0.038)	(0.073)
pct_{im}^{SC}	0.104***	-0.168***	0.013	0.050
	(0.017)	(0.032)	(0.037)	(0.060)
imp_{im}^{01}	0.059***	0.136***	0.012	0.045***
	(0.006)	(0.018)	(0.010)	(0.016)
Top Department Fixed Effects	Y	Y	Y	Y
Age Fixed Effects	Y	Y	Y	Y
Market Controls	Y	Y	Y	Y
R2	0.122	0.094	0.064	0.049
Observations	488,000	259,000	87,000	56,000

Notes: Exit is an indicator of whether store i exits prior to 2007. Pct change is the percent change in sales between 2002 and 2007. $\Delta Z_{im}^{2002-2007}$ is predicted the predicted change import exposure of store i using exports from China. $\Delta X_{im}^{2002-2007}$ is the change in direct import exposure of the store between 2002 and 2007. X_{im}^{2002} is the level of exposure in 2002. pct_{im}^{w} is the exposure of store i to firms of type $w \in$ {Large, General Merchandiser, Small-Chain}. imp_{im}^{01} is an indicator that a store imports at least 1 percent of its sales directly. Regressions include fixed effects for top department of each store and for the age of the store as well as controls for the log of store sales in 2002 and the change in commuting zone population between 2002 and 2007. Standard errors are clustered at the commuting zone-top department-level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The observation counts are rounded to the nearest thousand.

Table B.3: Change in Exposure to Large Retailers

	Single-Unit		Small Chain	
Dependent variable is:	Exit	Pct Change	Exit	Pct Change
$\Delta pct_{im}^{L,2002-2007}$	0.322***	-0.296***	0.374***	-0.289***
	(0.028)	(0.027)	(0.037)	(0.054)
pct_{im}^{L}	0.144***	-0.218***	0.161***	-0.232***
	(0.019)	(0.021)	(0.026)	(0.038)
pct_{im}^{SC}	0.081***	-0.092***	0.011	0.048
	(0.019)	(0.030)	(0.035)	(0.053)
Log sales	-0.010***	-0.108***	-0.079***	-0.074***
	(0.001)	(0.003)	(0.002)	(0.006)
Top Department Fixed Effects	Y	Y	Y	Y
Age Fixed Effects	Y	Y	Y	Y
Market Controls	Y	Y	Y	Y
R2	0.122	0.094	0.065	0.052
Observations	488,000	259,000	87,000	56,000

Notes: Exit is an indicator of whether store i exits prior to 2007. Pct change is the percent change in sales between 2002 and 2007. $\Delta pct_{im}^{L,2002-2007}$ is the change in exposure to large stores between 2002 and 2007. pct_{im}^{w} is the exposure of store i to firms of type $w \in \{\text{Large, Small-Chain}\}$. Regressions include fixed effects for top department of each store and for the age of the store as well as controls for the log of store sales in 2002 and the change in commuting zone population between 2002 and 2007. Standard errors are clustered at the commuting zone-top department-level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The observation counts are rounded to the nearest thousand.

Table B.4: Change in Exposure to General Merchandisers

	Single-Unit		Small Chain	
Dependent variable is:	Exit	Pct Change	Exit	Pct Change
$\Delta pct_{im}^{GM,2002-2007}$	0.257***	-0.370***	0.017	-0.255***
	(0.024)	(0.030)	(0.041)	(0.059)
pct_{im}^{GM}	-0.058***	-0.246***	-0.139***	-0.022
	(0.017)	(0.026)	(0.031)	(0.048)
pct_{im}^{SC}	0.081***	-0.026	-0.077***	0.192***
	(0.019)	(0.025)	(0.029)	(0.044)
Log sales	-0.100***	-0.110***	-0.081***	-0.073***
	(0.000)	(0.003)	(0.002)	(0.006)
Top Department Fixed Effects	Y	Y	Y	Y
Age Fixed Effects	Y	Y	Y	Y
Market Controls	Y	Y	Y	Y
R2	0.122	0.094	0.065	0.051
Observations	488,000	259,000	87,000	56,000

Notes: Exit is an indicator of whether store i exits prior to 2007. Pct change is the percent change in sales between 2002 and 2007. $\Delta pct_{im}^{GM,2002-2007}$ is the change in exposure to general merchandise stores between 2002 and 2007. pct_{im}^{w} is the exposure of store i to firms of type $w \in$ {General Merchandisers, Small-Chain}. Regressions include fixed effects for top department of each store and for the age of the store as well as controls for the log of store sales in 2002 and the change in commuting zone population between 2002 and 2007. Standard errors are clustered at the commuting zone-top department-level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The observation counts are rounded to the nearest thousand.